

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:21 ; Search time 218.82 seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-3
Perfect score: 25
Sequence: 1 gttgctgactctgctgttgaagcg 25

Scoring table: OLIGO_MUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2-6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2-6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2-6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2-6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2-6/ptodata/2/ina/PCBUS.COMB.seq.*
6: /cgn2-6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7 Sequence 7, Appl
2	25	100.0	1041	2	US-08-846-338-7 Sequence 7, Appl
3	25	100.0	5872	3	US-08-411-768B-1 Sequence 1, Appl
4	25	100.0	5872	3	US-08-411-768B-6 Sequence 6, Appl
5	16	64.0	5000	3	US-09-104-070-1 Sequence 1, Appl
6	15	60.0	708	4	US-08-998-416-997 Sequence 9, App
7	14	56.0	384	2	US-08-820-170A-8 Sequence 8, Appl
8	14	56.0	384	2	US-09-055-699-8 Sequence 8, Appl
9	14	56.0	384	2	US-09-273-565-8 Sequence 8, Appl
10	14	56.0	735	5	PCT-US94-14106-54 Sequence 54, Appl
11	14	56.0	1493	2	US-08-820-170A-9 Sequence 9, Appl
12	14	56.0	1493	3	US-09-055-699-9 Sequence 9, Appl
13	14	56.0	1493	4	US-09-273-565-9 Sequence 9, Appl
14	14	56.0	1848	1	US-08-447-422-15 Sequence 15, Appl
15	14	56.0	3243	2	US-08-239-276-6 Sequence 6, Appl
16	14	56.0	3243	2	US-08-468-577B-6 Sequence 6, Appl
17	14	56.0	3243	3	US-08-468-577B-6 Sequence 6, Appl
18	14	56.0	3282	1	US-08-072-574-11 Sequence 11, Appl
19	14	56.0	3282	1	US-08-486-270-11 Sequence 11, Appl
20	14	56.0	3282	3	US-08-367-264-11 Sequence 11, Appl
21	14	56.0	4085	1	US-08-072-574-7 Sequence 7, Appl
22	14	56.0	4085	1	US-08-486-270-7 Sequence 7, Appl
23	14	56.0	4085	3	US-08-367-264-7 Sequence 7, Appl
24	14	56.0	4181	1	US-08-072-574-9 Sequence 9, Appl
25	14	56.0	4181	1	US-08-486-270-9 Sequence 9, Appl
26	14	56.0	4181	3	US-08-367-264-9 Sequence 9, Appl
27	14	56.0	4207	4	US-08-660-148-1 Sequence 1, Appl

C 28	14	56.0	4207	4	US-08-660-148-3	Sequence 3, Appl
C 29	14	56.0	4303	4	US-08-660-148-4	Sequence 4, Appl
C 30	14	56.0	4303	4	US-08-660-148-6	Sequence 6, Appl
C 31	14	56.0	4566	2	US-08-465-976A-1	Sequence 1, Appl
C 32	14	56.0	4566	2	US-08-982-412-1	Sequence 1, Appl
C 33	14	56.0	5319	1	US-08-169-927-1	Sequence 1, Appl
C 34	14	56.0	7326	1	US-08-194-468-1	Sequence 1, Appl
C 35	14	56.0	7344	3	US-08-961-739-1	Sequence 1, Appl
C 36	14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appl
C 37	14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appl
C 38	14	56.0	12884	2	US-08-592-874-1	Sequence 2, Appl
C 39	14	56.0	28804	3	US-09-096-942-2	Sequence 2, Appl
C 40	14	56.0	28804	3	US-09-096-867-2	Sequence 2, Appl
C 41	13	52.0	749	4	US-08-998-416-727	Sequence 727, App
C 42	13	52.0	1161	5	US-08-153-848-31	Sequence 31, Appl
C 43	13	52.0	1161	5	PCT-US93-11153-31	Sequence 31, Appl
C 44	13	52.0	1242	3	US-08-252-966B-13	Sequence 13, Appl
C 45	13	52.0	1729	3	US-09-188-930-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7


```

? LOCATION: 5098..5574
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 5098
? OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
? OTHER INFORMATION: /product= "protein"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ORF1"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 45..49
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "promoter plac"
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 23..28
? OTHER INFORMATION: /standard_name= "promoter plac"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 105..119
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "biob RBS no.9"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 2284..2297
? OTHER INFORMATION: /standard_name= "bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3742..3752
? OTHER INFORMATION: /standard_name= "bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 5088..5100
? OTHER INFORMATION: /standard_name= "ORF1 RBS"
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 5583..5644
? OTHER INFORMATION: /standard_name= "rho-independent
? OTHER INFORMATION: transcriptional terminator"
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5583..5605
? FEATURE:
? NAME/KEY: promoter
? LOCATION: 1..96
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "promoter plac"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
? FILING DATE: 26-AUG-1986
? PUBLICATION DATE: 07-APR-1993
? US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtctgcatctgctgtttgaagc 25
Db 173 GTTGTGATCTGCTGTTTGAAGCG 197

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brunnaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1154..2308
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 1154
? OTHER INFORMATION: /ec_number= 2.3.1.47
? OTHER INFORMATION: /product= "KAPA synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioc"
? OTHER INFORMATION: /number= 2
? OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3043..3753
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 3043
? OTHER INFORMATION: /ec_number= 6.3.3.3
? OTHER INFORMATION: /product= "DTB synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioc"
? OTHER INFORMATION: /number= 4
? OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 1141..1156
? OTHER INFORMATION: /standard_name= "bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3030..3045
? OTHER INFORMATION: /standard_name= "bioc RBS"
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
```

FILED DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgcatctgctgtttgaagc 25
|||||
Db 173 GTTCTGATCTGCTTTGAAGC 197

RESULT 5

US-09-104-070-1

Sequence 1, Application US/09104070
Patent No. 6153741

GENERAL INFORMATION:

APPLICANT: Richards, Eric J.
TITLE OF INVENTION: DNA Methylation Gene from Plants
FILE REFERENCE: Wash U CI-0014 US
CURRENT APPLICATION NUMBER: US/09/104,070
EARLIER FILING DATE: 1998-06-24
EARLIER APPLICATION NUMBER: US 60/083,612
NUMBER OF SEQ. ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ. ID NO 1
LENGTH: 5000
TYPE: DNA
ORGANISM: Arabidopsis thaliana

FEATURE:
NAME/KEY: CDS

LOCATION: (782)...(1252)

NAME/KEY: CDS

LOCATION: (1354)...(1440)

NAME/KEY: CDS

LOCATION: (1549)...(1895)

NAME/KEY: CDS

LOCATION: (1976)...(2165)

NAME/KEY: CDS

LOCATION: (2251)...(2437)

NAME/KEY: CDS

LOCATION: (2559)...(2629)

NAME/KEY: CDS

LOCATION: (2703)...(2892)

NAME/KEY: CDS

LOCATION: (2975)...(3070)

NAME/KEY: CDS

LOCATION: (3148)...(3242)

NAME/KEY: CDS

LOCATION: (3317)...(3436)

NAME/KEY: CDS

LOCATION: (3540)...(3659)

NAME/KEY: CDS

LOCATION: (3745)...(3843)

NAME/KEY: CDS

LOCATION: (3934)...(4038)

NAME/KEY: CDS

LOCATION: (4130)...(4354)

NAME/KEY: CDS

LOCATION: (4826)...(4755)

OTHER INFORMATION: t-RNA-glu coding region

FEATURE:

NAME/KEY: mutation

LOCATION: (785)...(786)

OTHER INFORMATION: som8 rearrangement, deletion of G at 785 and

FEATURE: insertion of 83 bp between 785 and 786

NAME/KEY: mutation

LOCATION: (3243)...(3243)

OTHER INFORMATION: ddm1-2 base pair substitution of G to A
US-09-104-070-1

Query Match 64.0%; Score 16; DB 3; Length 5000;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcggatctgctgt 18
|||||
Db 1132 tgcggatctgctgt 1147

RESULT 6

US-08-998-416-997/c

Sequence 997, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPTI
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 997:
SEQUENCE CHARACTERISTICS:

LENGTH: 708 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:

ORGANISM: PAG1612RP
US-08-998-416-997

Query Match 60.0%; Score 15; DB 4; Length 708;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcggatctgctgt 17
|||||
Db 470 TGCTGATCTGCTGT 456

RESULT 7
US-08-820-170A-8/C
Sequence 8, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-08-820-170A-8

Query Match 56.0%; Score 14; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
|||||

Db 350 TGGATCTGCTGTTT 337

RESULT 8
US-09-055-699-8/C
Sequence 8, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/820,170
APPLICATION NUMBER:
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-09-055-699-8

Query Match 56.0%; Score 14; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
|||||

Db 350 TGGATCTGCTGTTT 337

RESULT 9
US-09-273-565-8/C
Sequence 8, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
SKELETAL MUSCLE-SPECIFIC UBQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 8
LENGTH: 384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-273-565-8

Query Match 56.0%; Score 14; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
|||||

Db 350 TGGATCTGCTGTTT 337

RESULT 10
PCT-US94-14106-54

```
; Sequence 54, Application PC/TUS9414106
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for Generating Specific Antibodies
; NUMBER OF SEQUENCES: 61
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14106
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..735
; PCT-US94-14106-54

Query Match          56.0%; Score 14; DB 5; Length 735;
Best Local Similarity 100.0%; Pred No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tctgtatctgctg 16
Db 465 TCGTGATCTGCTG 478

RESULT 11
US-08-820-170A-9/c
; Sequence 9, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-08-820-170A-9
```

```
Query Match          56.0%; Score 14; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 tggatctgctgtt 19
Db 444 TGGATCTGCTGTTT 431
```

```
RESULT 12
US-09-055-699-9/c
; Sequence 9, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-09-055-699-9
```

Query Match 56.0%; Score 14; DB 3; Length 1493;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgtt 19
|||||
DB 444 TGGATCTGCTGTT 431

RESULT 13
US-09-273-565-9/c
; Sequence 9, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIMURA, TSUTOMU
; APPLICANT: MATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ. ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ. ID NO. 9
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95)..(478)
US-09-273-565-9

Query Match 56.0%; Score 14; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgtt 19
|||||
DB 444 TGGATCTGCTGTT 431

RESULT 14
US-08-447-422-15/c
; Sequence 15, Application US/08447422
; Patent No. 5686579
; GENERAL INFORMATION:
; APPLICANT: SHAMI, Ezekiel Y.
; APPLICANT: ROTHSTEIN, Asael
; APPLICANT: RAMDEESINGH, Mohanb
; TITLE OF INVENTION: Use of Antibody/Antigen Interactions To
; TITLE OF INVENTION: Protect or Modulate Biological Activity
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,422
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,410
; FILING DATE: 22-JUN-1993
; APPLICATION NUMBER: US 07/938,505
; FILING DATE: 31-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/205,748
; FILING DATE: 21-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 17923/102 HYLI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ. ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
US-08-447-422-15

Query Match 56.0%; Score 14; DB 1; Length 1848;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgctgctg 16
|||||
DB 1829 TCGTGTGCTGCTG 1816

RESULT 15
US-08-239-276-6/c
; Sequence 6, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rablin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-239-276-6

Query Match 56.0%; Score 14; DB 2; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactctgctg 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 16
US-08-468-579B-6/c
Sequence 6, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
NUMBER OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-579B-6

Query Match 56.0%; Score 14; DB 2; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactctgctg 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 17
US-08-468-577B-6/c
Sequence 6, Application US/08468577B
Patent No. 6001804
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
NUMBER OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-577B-6

Query Match 56.0%; Score 14; DB 3; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 18
US-08-072-574-11/c
; Sequence 11, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NOCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,574
; FILING DATE: 19930604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product= "HUMAN MGLUR5C"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with truncated 3' end."
US-08-072-574-11

Query Match 56.0%; Score 14; DB 1; Length 3282;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 19
US-08-486-270-11/c
; Sequence 11, Application US/08486270

; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NOCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-9392
; TELEFAX: 619-546-4737
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product= "HUMAN MGLUR5C"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with truncated 3' end."
US-08-486-270-11

Query Match 56.0%; Score 14; DB 1; Length 3282;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 20
US-08-367-264-11/c
; Sequence 11, Application US/08367264
; Patent No. 6001581
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,

;; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;; NUMBER OF SEQUENCES: 13
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
;; STREET: 444 South Flower Street, Suite 2000
;; CITY: Los Angeles
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 90071
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/367,264
;; FILING DATE: 02-JUN-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/072,574
;; FILING DATE: 04-JUN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reiter, Stephen E.
;; REGISTRATION NUMBER: 31,192
;; REFERENCE/DOCKET NUMBER: P41 9772
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-546-4737
;; TELEFAX: 619-546-9392
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3282 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: both
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 370..3003
;; OTHER INFORMATION: /product="HUMAN MGLUR5C"
;; OTHER INFORMATION: /note="Variant of MGLUR5A with truncated 3' end."
US-08-367-264-11

Query Match 56.0%; Score 14; DB 3; Length 3282;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctgactcgtc 16
|||||
DB 2950 tgctgactcgtc 2937

RESULT 21
US-08-072-574-7/c
; Sequence 7, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/072,574
;; FILING DATE: 19930604
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reiter, Stephen E.
;; REGISTRATION NUMBER: 31,192
;; REFERENCE/DOCKET NUMBER: P41 9383
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 213-622-7700
;; TELEFAX: 213-489-4210
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4085 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: both
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 370..3912
;; OTHER INFORMATION: /product="HUMAN MGLUR5A"
US-08-072-574-7

Query Match 56.0%; Score 14; DB 1; Length 4085;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctgactcgtc 16
|||||
DB 2950 tgctgactcgtc 2937

RESULT 22
US-08-486-270-7/c
; Sequence 7, Application US/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4085 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3912
OTHER INFORMATION: /product= "HUMAN MGLUR5A"
US-08-486-270-7

Query Match 56.0%; Score 14; DB 1; Length 4085;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtc 16
|||||
DB 2950 TCCTGATCTCTG 2937

RESULT 23
US-08-367-264-7/C
Sequence 7, Application US/08367264
Patent No. 6001581
GENERAL INFORMATION:
APPLICANT: Daggett, Lorie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESS: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4085 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both

MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3912
OTHER INFORMATION: /product= "HUMAN MGLUR5A"
US-08-367-264-7

Query Match 56.0%; Score 14; DB 3; Length 4085;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtc 16
|||||
DB 2950 TCCTGATCTCTG 2937

RESULT 24
US-08-072-574-9/C
Sequence 9, Application US/08072574
Patent No. 5521297
GENERAL INFORMATION:
APPLICANT: Daggett, Lorie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESS: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 00719
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,574
FILING DATE: 19930604
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGLUR5B"
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
insertion between nucleotides 2999 and 3000."
US-08-072-574-9

Query Match 56.0%; Score 14; DB 1; Length 4181;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 tgcgtgactcgtc 16

Db 2950 TCGTGATCTGCTG 2937

RESULT 25

US-08-486-270-9/c
Sequence 9, Application US/08486270
Patent No. 5807689

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrise
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.

TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,270
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGUR5B"
OTHER INFORMATION: /note= "Variant of MGUR5A with 96 base pair
insertion between nucleotides 2998 and 2999."

US-08-486-270-9

Query Match

Best Local Similarity 56.0%; Score 14; DB 1; Length 4181;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2950 TCGTGATCTGCTG 2937

RESULT 26

US-08-367-264-9/c
Sequence 9, Application US/08367264
Patent No. 6001581

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrise
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.

TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGUR5B"
OTHER INFORMATION: /note= "Variant of MGUR5A with 96 base pair
insertion between nucleotides 2998 and 2999."

US-08-367-264-9

Query Match

Best Local Similarity 56.0%; Score 14; DB 3; Length 4181;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2950 TCGTGATCTGCTG 2937

RESULT 27

US-08-660-148-1/c
Sequence 1, Application US/08660148
Patent No. 6211353

GENERAL INFORMATION:

APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
APPLICANT: Hess, Stephen D.

TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /product= "HUMAN MGUR5B"
OTHER INFORMATION: /note= "Variant of MGUR5A with 96 base pair
insertion between nucleotides 2998 and 2999."

US-08-660-148-1/c

Query Match

Best Local Similarity 56.0%; Score 14; DB 3; Length 4181;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2950 TCGTGATCTGCTG 2937

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Eli Lilly and Company
3 STREET: Lilly Corporate Center
4 CITY: Indianapolis
5 STATE: Indiana
6 COUNTRY: United States of America
7 ZIP: 46285
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentin Release #1.0, Version #1.25
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/660,148
16
17 FILING DATE:
18 CLASSIFICATION:
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US/08/282,853
21 FILING DATE:
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Gaylo, Paul J.
24 REGISTRATION NUMBER: 36,808
25 REFERENCE/DOCKET NUMBER: X-9419
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (317) 276-0756
28 TELEFAX: (317) 276-3861
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 4207 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: CDNA
36 FEATURE:
37 NAME/KEY: CDS
38 LOCATION: 460..3999
39 US-08-660-148-1
40
41 Query Match 56.0%; Score 14; DB 4; Length 4207;
42 Best Local Similarity 100.0%; Pred. No. 21;
43 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
44
45 QY 3 tgcgtgacatctgctg 16
46 Db 3040 TGCCTGATCTGCTG 3027
47
48 RESULT 28
49 US-08-660-148-3/c
50 Sequence 3, Application US/08660148
51 Patent No. 6211353
52 GENERAL INFORMATION:
53 APPLICANT: Burnett, J. P.
54 APPLICANT: Mayne, Nancy G.
55 APPLICANT: Sharp, Robert L.
56 APPLICANT: Snyder, Yvonne M.
57 TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
58 TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
59 NUMBER OF SEQUENCES: 6
60 CORRESPONDENCE ADDRESS:
61 ADDRESSEE: Eli Lilly and Company
62 STREET: Lilly Corporate Center
63 CITY: Indianapolis
64 STATE: Indiana
65 COUNTRY: United States of America
66 ZIP: 46285
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS
71 SOFTWARE: Patentin Release #1.0, Version #1.25
72 CURRENT APPLICATION DATA:
73 APPLICATION NUMBER: US/08/660,148
74 FILING DATE:
75 CLASSIFICATION:
76 PRIOR APPLICATION DATA:
77 APPLICATION NUMBER: US/08/282,853
78 FILING DATE:
79 ATTORNEY/AGENT INFORMATION:
80 NAME: Gaylo, Paul J.
81 REGISTRATION NUMBER: 36,808
82 REFERENCE/DOCKET NUMBER: X-9419
83 TELECOMMUNICATION INFORMATION:
84 TELEPHONE: (317) 276-0756
85 TELEFAX: (317) 276-3861
86 INFORMATION FOR SEQ ID NO: 4:
87 SEQUENCE CHARACTERISTICS:
88 LENGTH: 4303 base pairs
89 TYPE: nucleic acid
90 STRANDEDNESS: single
91 TOPOLOGY: linear
92 MOLECULE TYPE: mRNA
93 US-08-660-148-3

1 APPLICATION NUMBER: US/08/660,148
2 FILING DATE:
3 CLASSIFICATION:
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: US/08/282,853
6 FILING DATE:
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Gaylo, Paul J.
9 REGISTRATION NUMBER: 36,808
10 REFERENCE/DOCKET NUMBER: X-9419
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (317) 276-0756
13 TELEFAX: (317) 276-3861
14 INFORMATION FOR SEQ ID NO: 3:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 4207 base pairs
17 TYPE: nucleic acid
18 STRANDEDNESS: single
19 TOPOLOGY: linear
20 MOLECULE TYPE: mRNA
21 US-08-660-148-3
22
23 Query Match 56.0%; Score 14; DB 4; Length 4207;
24 Best Local Similarity 100.0%; Pred. No. 21;
25 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
26
27 QY 3 tgcgtgacatctgctg 16
28 Db 3040 TGCCTGATCTGCTG 3027
29
30 RESULT 29
31 US-08-660-148-4/c
32 Sequence 4, Application US/08660148
33 Patent No. 6211353
34 GENERAL INFORMATION:
35 APPLICANT: Burnett, J. P.
36 APPLICANT: Mayne, Nancy G.
37 APPLICANT: Sharp, Robert L.
38 APPLICANT: Snyder, Yvonne M.
39 TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
40 TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
41 NUMBER OF SEQUENCES: 6
42 CORRESPONDENCE ADDRESS:
43 ADDRESSEE: Eli Lilly and Company
44 STREET: Lilly Corporate Center
45 CITY: Indianapolis
46 STATE: Indiana
47 COUNTRY: United States of America
48 ZIP: 46285
49 COMPUTER READABLE FORM:
50 MEDIUM TYPE: Floppy disk
51 COMPUTER: IBM PC compatible
52 OPERATING SYSTEM: PC-DOS/MS-DOS
53 SOFTWARE: Patentin Release #1.0, Version #1.25
54 CURRENT APPLICATION DATA:
55 APPLICATION NUMBER: US/08/660,148
56 FILING DATE:
57 CLASSIFICATION:
58 PRIOR APPLICATION DATA:
59 APPLICATION NUMBER: US/08/282,853
60 FILING DATE:
61 ATTORNEY/AGENT INFORMATION:
62 NAME: Gaylo, Paul J.
63 REGISTRATION NUMBER: 36,808
64 REFERENCE/DOCKET NUMBER: X-9419
65 TELECOMMUNICATION INFORMATION:
66 TELEPHONE: (317) 276-0756
67 TELEFAX: (317) 276-3861
68 INFORMATION FOR SEQ ID NO: 4:
69 SEQUENCE CHARACTERISTICS:
70 LENGTH: 4303 base pairs
71 TYPE: nucleic acid
72 STRANDEDNESS: single
73 TOPOLOGY: linear
74 MOLECULE TYPE: mRNA
75 US-08-660-148-3

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 460..4095
US-08-660-148-4

Query Match

Best Local Similarity 56.0%; Score 14; DB 4; Length 4303;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3040 TGCTGATCTGCTG 3027

RESULT 30

US-08-660-148-6/c
Sequence 6, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
NUMBER OF SEQUENCES: 6
RELATED NUCLEIC ACID COMPOUNDS
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-6

Query Match

Best Local Similarity 56.0%; Score 14; DB 4; Length 4303;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 tctgtgactctgtg 16 ,
|||||

Db 3040 TGCTGATCTGCTG 3027

RESULT 31

US-08-465-976A-1/c
Sequence 1, Application US/08465976A
Patent No. 5869632
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN GIFFILLAN, CECCHI
ADDRESSEE: STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NJ
COUNTRY: US
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,976A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY F
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-444
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 994-1700
TELEFAX: (201) 994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 212..2863
US-08-465-976A-1

Query Match

Best Local Similarity 56.0%; Score 14; DB 2; Length 4566;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

5 ctgtatctgtgtt 18
Db 4182 CTGATCTGCTGTT 4169
|||||

RESULT 32

US-08-982-412-1/c
Sequence 1, Application US/08982412
Patent No. 5958729
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.

STREET: 9410 KEY WEST AVENUE
 CITY: ROCKVILLE,
 STATE: MD
 COUNTRY: US
 ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/982,412
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: BROOKES, ANDERS A
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PF181PCT2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4566 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 212..2863
 US-08-982-412-1

Query Match 56.0%; Score 14; DB 2; Length 4566;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctagatcgcgtt 18
 |||||||
 Db 4182 CTGCATCTGCTGTT 4169

RESULT 33
 US-08-169-927-1
 ; Sequence 1, Application US/08169927
 ; Patent No. 5783441
 ; GENERAL INFORMATION:
 ; APPLICANT: Carl, Mitchell
 ; APPLICANT: Dobson, Michael E.
 ; APPLICANT: Ching, Wei Mei
 ; TITLE OF INVENTION: Gene and protein Applicable to the
 ; TITLE OF INVENTION: Preparation of vaccines for Rickettsia prowazekii and
 ; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Counsel, Naval Medical R & D Command
 ; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
 ; CITY: Bethesda
 ; STATE: MD
 ; COUNTRY: USA
 ; ZIP: 20889-5606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/169,927
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/742,128
 FILING DATE: 08/09/91
 ATTORNEY/AGENT INFORMATION:
 NAME: Spevack, A. David
 REGISTRATION NUMBER: 24,743
 REFERENCE/DOCKET NUMBER: 75,976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 295-6759
 TELEFAX: (301) 295-1022
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5319 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE: Rickettsia prowazekii
 ORGANISM: Rickettsia
 STRAIN: Breinl

FEATURE:
 NAME/KEY: -35_signal
 LOCATION: 340..345
 FEATURE:
 NAME/KEY: -10_signal
 LOCATION: 363..368
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 391..5226
 FEATURE:
 NAME/KEY: RBS
 LOCATION: 379..386
 FEATURE:
 NAME/KEY: stem_loop
 LOCATION: 5270..5306
 PUBLICATION INFORMATION:
 AUTHORS: Carl, M.
 AUTHORS: Dobson, M. E.
 AUTHORS: Ching, W. M.
 AUTHORS: Dasch, G. A.
 TITLE: Characterization of the gene encoding the
 TITLE: protective s-layer protein of Rickettsia
 TITLE: prowazekii; presence of a truncated identical
 TITLE: homolog in rickettsia typhi
 JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
 DATE: 1990
 RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
 US-08-169-927-1

Query Match 56.0%; Score 14; DB 1; Length 5319;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttagcgcgttc 14
 |||||||
 Db 4321 GTTCCTGATCTGC 4334

RESULT 34
 US-08-194-468-1/c
 ; Sequence 1, Application US/08194468
 ; Patent No. 5750336
 ; GENERAL INFORMATION:
 ; APPLICANT: Montminy, Marc R.
 ; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
 ; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
 ; TITLE OF INVENTION: RESPONSIVE GENES
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/194,468

FILING DATE: 10-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: P41 9672

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619)-546-4737

TELEFAX: (619)-546-9392

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7326 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..7323

US-08-194-468-1

Query Match 56.0%; Score 14; DB 1; Length 7326;
Best Local Similarity 100.0%; Pred. No. 21;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tcttgatctgctg 16

Db 5573 TCGTGATCTGCTG 5560

RESULT 35

US-08-961-739-1/C

Sequence 1, Application US/08961739A

Patent No. 6083583

GENERAL INFORMATION:

APPLICANT: Montminy, Marc R.

TITLE OF INVENTION: Methods for Treating Diabetes Mellitus

FILE REFERENCE: SALK1650-1

CURRENT APPLICATION NUMBER: US/08/961,739A

EARLIER FILING DATE: 1997-10-31

EARLIER APPLICATION NUMBER: US 194,468

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 1

LENGTH: 7326

TYPE: DNA

ORGANISM: Mus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(7326)

NAME/KEY: misc feature

LOCATION: (1)...(7326)

OTHER INFORMATION: n = A,T,C or G

US-08-961-739-1

Query Match 56.0%; Score 14; DB 3; Length 7344;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tcttgatctgctg 16

Db 5573 TCGTGATCTGCTG 5560

RESULT 36

US-08-876-991-1

Sequence 1, Application US/08876991

Patent No. 5925360

GENERAL INFORMATION:

APPLICANT: Gregor Meyers, Tillmann R menapf,

TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Organon Teknika Corporation

STREET: 1330-A Piccard Drive

CITY: Rockville

STATE: Maryland

COUNTRY: U.S.A.

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/876,991

FILING DATE: 16-JUN-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/747,577

FILING DATE: US/08/650,584

APPLICATION NUMBER: US/08/469,702

FILING DATE: US/08/123,596

APPLICATION NUMBER: 07/797,554

FILING DATE: 22-NOV-1991

APPLICATION NUMBER: US 07/494,991

FILING DATE: 16-MAR-1990

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: William M. Blackstone

REGISTRATION NUMBER: 29,772

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 12284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Hog cholera virus

STRAIN: Alfort

CELL LINE: PK 15 and 38A1D

FEATURE:

NAME/KEY: CDS

LOCATION: 364..12060

OTHER INFORMATION: /label= 435_kDa_protein

FEATURE:

NAME/KEY: primer bind

LOCATION: complement (2587..2619)

OTHER INFORMATION: /label= primer_1

FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
|||||
Db 2253 TCTGCTGTGGAAG 2266

RESULT 37
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESSEE: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D

FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-09-059-853-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
|||||
Db 2253 TCTGCTGTGGAAG 2266

RESULT 38
US-08-592-874-1/c
Sequence 1, Application US/08592874
Patent No. 5854034
GENERAL INFORMATION:
APPLICANT: POLLOCK, THOMAS J.
APPLICANT: YAMAZAKI, MOTOHIDE
APPLICANT: THORNE, LINDA
APPLICANT: MIKOLAJCZAK, MARCIA
APPLICANT: ARMENTROUT, RICHARD W.
TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: JULES E. GOLDBERG
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,874
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/377,440
FILING DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 28804 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
US-08-592-874-1

Query Match
Best Local Similarity 100.0%; Score 14; DB 2; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctggatctgctg 16
Db 14585 TGCTGGATCTGCTG 14572

RESULT 39
US-09-096-942-2/c
Sequence 2, Application US/09096942
Patent No. 6027925
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteria*
TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,942
EARLIER FILING DATE: 1998-06-12
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 28804
TYPE: DNA
ORGANISM: *Sphingomonas* sp. S88
US-09-096-942-2

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctggatctgctg 16
Db 14585 TGCTGGATCTGCTG 14572

RESULT 40
US-09-096-867-2/c
Sequence 2, Application US/09096867
Patent No. 6030817
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteria*
TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,867
EARLIER FILING DATE: 1998-06-11
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2

LENGTH: 28804
TYPE: DNA
ORGANISM: *Sphingomonas* sp. S88
US-09-096-867-2

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctggatctgctg 16
Db 14585 TGCTGGATCTGCTG 14572

RESULT 41
US-08-998-416-727
Sequence 727, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF *ASHBYA GOSSYPII*
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 727:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1475RP
US-08-998-416-727

Query Match
Best Local Similarity 100.0%; Score 13; DB 4; Length 749;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tgctgtttgaagc 24

Db 96 TGCTGTTGAAGC 108

RESULT 42

US-08-153-848-31
Sequence 31, Application US/08153848
Patent No. 5759804

GENERAL INFORMATION:

APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.

TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell

STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848

CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:

LENGTH: 1161 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: exon

LOCATION: 7..80

FEATURE:
NAME/KEY: CDS

LOCATION: 94..1158

US-08-153-848-31

Query Match

Best Local Similarity 52.0%; Score 13; DB 1; length 1161;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gatctgctgttg 20

Db 322 GATCTGCTGTTTG 334

RESULT 43

PCT-US93-11153-31

Sequence 31, Application PC/TUS9311153

GENERAL INFORMATION:

APPLICANT: Godiska, Ronald

APPLICANT: Gray, Patrick W.

APPLICANT: Schweikart, Vicki L.

TITLE OF INVENTION: Novel Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell

STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Noland, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:

LENGTH: 1161 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: exon

LOCATION: 7..80

FEATURE:
NAME/KEY: CDS

LOCATION: 94..1158

PCT-US93-11153-31

Query Match

Best Local Similarity 52.0%; Score 13; DB 5; length 1161;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gatctgctgttg 20

Db 322 GATCTGCTGTTTG 334

RESULT 44

US-08-252-966B-13/c

Sequence 13, Application US/08252966B

Patent No. 5624818

GENERAL INFORMATION:

APPLICANT: Eisenman, Robert N.

APPLICANT: Huth, Peter J.

APPLICANT: Ayer, Donald E.

TITLE OF INVENTION: Regulatory proteins that dimerize with

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC

STREET: 1420 Fifth Ave., Suite 2800

CITY: Seattle

STATE: Washington

COUNTRY: USA

Search completed: October 9, 2001, 15:55:26
 Job time: 13202 sec

ZIP: 98101-2347
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/252,966B
 FILING DATE: 01-JUN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Shelton, Dennis K.
 REGISTRATION NUMBER: 26,997.
 REFERENCE/DOCKET NUMBER: FHCRI7694
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-8100
 TELEFAX: (206) 224-0779
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1242 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 DESCRIPTION: clone 10; see figure 24
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Mus musculus
 US-08-252-966B-13

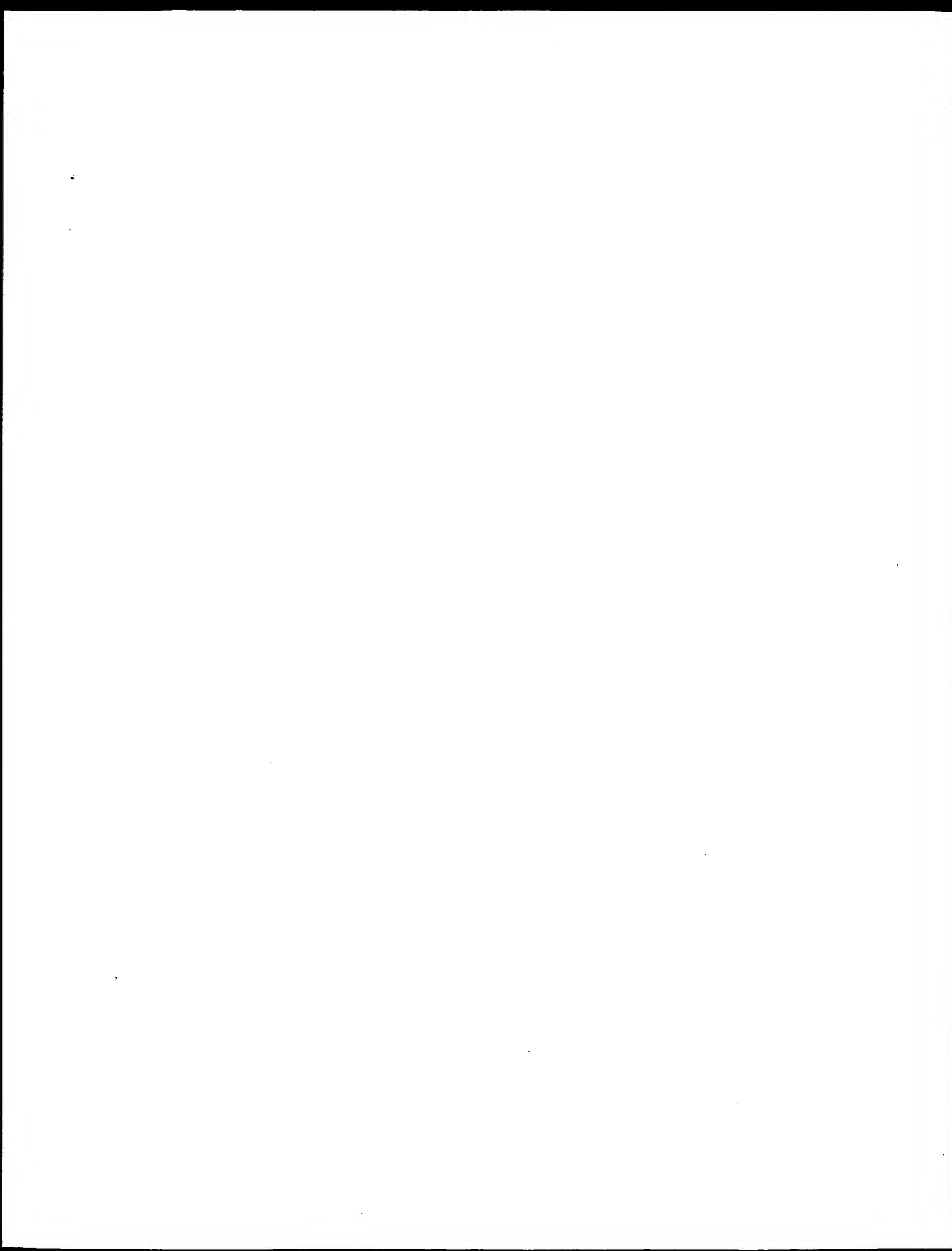
Query Match 52.0%; Score 13; DB 1; Length 1242;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ctgctgtttgaag 23
 |||||
 Db 525 CTGCTGTTCAG 513

RESULT 45
 US-09-188-930-68/c
 Sequence 68, Application US/09188930A
 Patent No. 6150502
 GENERAL INFORMATION:
 APPLICANT: Watson, James D.
 APPLICANT: Strachan, Lorna
 APPLICANT: Sleeman, Matthew
 APPLICANT: Onrust, Rene
 APPLICANT: Murison, James Greg
 TITLE OF INVENTION: Compositions Isolated From Skin Cells
 TITLE OF INVENTION: and Methods For Their Use
 FILE REFERENCE: 11000.1011c1
 CURRENT APPLICATION NUMBER: US/09/188,930A
 CURRENT FILING DATE: 1998-11-09
 NUMBER OF SEQ ID NOS: 348
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 68
 LENGTH: 1729
 TYPE: DNA
 ORGANISM: mouse
 US-09-188-930-68

Query Match 52.0%; Score 13; DB 3; Length 1729;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgagctgctgt 17
 |||||
 Db 1130 CTGAGCTGCTGT 1118



Wed Oct 10 07:45:47 2001

us-09-396-196f-3.ol1.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:23 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25
Sequence: 1 gttgctgactgtctgttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9 754579

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	76.0	888	221	CNS04KH2	AL294815 Tetraodon
2	18	72.0	403	163	BE095496	BE095496 UI-R-BU0
3	18	72.0	487	148	BF428015	BF428015 daa04g08.
4	17	68.0	546	236	A0955256	A0955256 Sheared D
5	17	68.0	547	32	AV718673	AV718673 AV718673
6	16	64.0	237	111	AM101291	AM101291 sq77b02.Y
7	16	64.0	247	162	BE023256	BE023256 sm80a02.Y
8	16	64.0	274	148	BF414722	BF414722 UI-R-BJ2
9	16	64.0	287	134	BB455974	BB455974 BB455974
10	16	64.0	287	171	BF917535	BF917535 113-07011
11	16	64.0	293	240	AZ214122	AZ214122 Sheared D
12	16	64.0	301	114	AM306120	AM306120 sc46b06.Y
13	16	64.0	328	138	BE659333	BE659333 GM700009A
14	16	64.0	342	115	AM423279	AM423279 sh65d11.Y
15	16	64.0	360	156	C61070	C61070 C61070 Yuj1
16	16	64.0	365	138	BE662000	BE662000 bs02d05.Y
17	16	64.0	422	24	A1736967	A1736967 sb35c02.Y
18	16	64.0	434	119	AM704839	AM704839 sk40b11.Y
19	16	64.0	436	18	A1293622	A1293622 LP06802.5
20	16	64.0	441	164	BE190258	BE190258 sc09f08.Y
21	16	64.0	470	105	AL373871	AL373871 MBB03D01
22	16	64.0	474	119	AM704996	AM704996 sk41c02.Y
23	16	64.0	481	105	AL379973	AL379973 MBB48E08
24	16	64.0	497	149	BF461669	BF461669 UI-M-CG0P
25	16	64.0	502	136	BE515828	BE515828 MHE0606.B
26	16	64.0	503	105	AL383090	AL383090 MRC01G02
27	16	64.0	517	105	AL373872	AL373872 MBB03D01
28	16	64.0	517	232	A0692054	A0692054 HS-5408_A
29	16	64.0	521	148	BF419551	BF419551 UI-R-CA0
30	16	64.0	522	102	AI854720	AI854720 UI-M-BH0
31	16	64.0	525	105	AL379974	AL379974 MBB48E08
32	16	64.0	531	115	AM395303	AM395303 sh46d09.Y
33	16	64.0	535	235	A0889100	A0889100 HS-2204_A
34	16	64.0	574	118	AM573807	AM573807 EST316398
35	16	64.0	575	143	BE999070	BE999070 EST430793
36	16	64.0	580	150	BF525150	BF525150 UI-R-AF0
37	16	64.0	591	120	AM734894	AM734894 sk76d03.Y
38	16	64.0	601	240	AZ267971	AZ267971 RPCT-23-1
39	16	64.0	609	113	AM267738	AM267738 EST305866
40	16	64.0	616	123	AM980891	AM980891 EST392044
41	16	64.0	618	155	BC585328	BC585328 EST487092
42	16	64.0	621	115	AM422111	AM422111 f157a12.Y
43	16	64.0	627	137	BE593142	BE593142 WSL_98_A0
44	16	64.0	632	244	AZ449344	AZ449344 IM0247704
45	16	64.0	640	16	A1108765	A1108765 GH08003.5

ALIGNMENTS

RESULT 1
CNS04KH2 888 bp DNA GSS 21 MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 11603 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL294815.1 GI:8033395
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 888)
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT

Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis
Unpublished
2 (bases 1 to 888)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 888)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES
source
Location/Qualifiers
1..888
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="11603"
/clone_1lb="G"
/note="Genoscope sequence ID : C0BGL16CB02LPI-end : T7"
BASE COUNT 129 a 276 c 276 g 187 t 20 others
ORIGIN

Query Match 76.0%; Score 19; DB 221; Length 888;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 tgcgtgacgcgtgtga 21
DB 331 TCGTGCATCTGCTGTGA 349

RESULT 2
BE095496/c 403 bp mRNA EST 12 JUN-2000
LOCUS Rattus norvegicus
DEFINITION UI-R-BU0-apa-c-11-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
UI-R-BU0-apa-c-11-0-UI 3', mRNA sequence.
ACCESSION BE095496
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 403)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msosares@iuii.weeg.uiowa.edu
Oligo-dt track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers

```

source
1. 403
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-800-apa-c-11-0-UI"
/clone_lib="UI-R-800"
/seq_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-800
library is a subcloned library derived from a mixture of
eye and ganglia tissues. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subcloning has
been previously described in (Bonaldi, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG:560=None found"

BASE COUNT      85 a      147 c      141 g      30 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 18; DB 163; Length 403;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtcgtt 18
|||||
Db 245 GTTCCTGACTCTGCTGTT 228

RESULT 3
BF428015 487 bp mRNA EST 30-MAR-2001
LOCUS daa04q08.x1 NICHD XGC Ovl Xenopus laevis cDNA clone IMAGE:4055247
DEFINITION 3' similar to TR:093478 O93478 KINESIN LIKE PROTEIN 3. ; mRNA
sequence.
ACCESSION BF428015
VERSION BF428015.1 GI:11439676
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
REFERENCE 1 (bases 1 to 487)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
, B., Gibbons,M., Harvey,N., Rittler,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 50.
Location/Qualifiers
1. 487
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4055247"
/clone_lib="NICHD XGC Ovl"
/sex="female"
/lab_host="DH10B (phage-resistant)"
/note="Organ ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.0 kb. Constructed by Life
Technologies."

```

```

BASE COUNT      138 a      101 c      102 g      146 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 18; DB 148; Length 487;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtcgtt 18
|||||
Db 249 GTTCCTGACTCTGCTGTT 266

RESULT 4
AG952526 546 bp DNA GSS 27-JAN-2000
LOCUS Sheared DNA-48019.TF Sheared DNA Trypanosoma brucei genomic clone
DEFINITION Sheared DNA-48019, DNA sequence.
ACCESSION AG952526
VERSION AG952526.1 GI:6775791
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
REFERENCE 1 (bases 1 to 546)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gull,S., Sub,E., Malek,J., Fujita,C.,
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Doneison,J.,
Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei
10.1 sheared DNA library
Unpublished (1999)
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@ligr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-forward
Class: Shotgun.
FEATURES
Location/Qualifiers
1. 546
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-48019"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + 1 method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."
BASE COUNT      192 a      161 c      127 g      66 t
ORIGIN

Query Match
Best Local Similarity 68.0%; Score 17; DB 236; Length 546;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtcgtt 17
|||||

```


Db 192 GTTCTGATCTGCTGT 176

RESULT 5
LOCUS AV718673 547 bp mRNA EST 16-OCT-2000
DEFINITION AV718673 GLC Homo sapiens cDNA clone GLCDA12 5', mRNA sequence.
ACCESSION AV718673
VERSION AV718673.1 GI:10815825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS Qian, B., Wu, T., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z., and Han, Z.
TITLE Homo sapiens cDNA GLC clones
JOURNAL Unpublished (2000)
CONTACT Zengqiang Han
CHINESE NATIONAL Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhongjiajiaqiao Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
TEL: 86-21-50801919 (ex.45)
FAX: 86-21-50801922
EMAIL: hanzq@hgc.sh.cn
This clone is available at CHGC in Shanghai.
LOCATION/Qualifiers
1. 547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCDA12"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 183 a 101 c 116 g 147 t

ORIGIN

Query Match 68.0%; Score 17; DB 32; Length 547;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttctgcatctgctgt 17
|||||

Db 473 GTTCTGATCTGCTGT 457

RESULT 6
LOCUS AM101291 237 bp mRNA EST 06-DEC-1999
DEFINITION sd77b02.y1 Gm-c1009 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1009-388 5', similar to TR:Q40380 ARABINOGLACTIN-PROTEIN PRECURSOR.; mRNA sequence.
ACCESSION AM101291
VERSION AM101291.1 GI:6071904
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 237)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erdelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wolfe, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

TITLE
JOURNAL
COMMENT
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco.
LOCATION/Qualifiers
1. 237
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1009-388"
/clone_lib="Gm-c1009"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II Xr; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from entire roots of 2-month-old 'Williams' plants that were greenhouse grown in 5-gallon pots. To suppress nodulation, Black Gold All-Purpose potting soil was supplemented with: 0.36g/L available phosphoric acid (P205), 20mg/L urea N, 0.16g/L S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.5mg/L Mn, 0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following nutrients in a slow-release form (Osmocote): 0.165g/L ammonia N, 0.185g/L nitrate N, 0.35g/L available phosphoric acid, and 0.35g/L soluble potash. No nodules were visible on the roots at harvest. Stragene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stragene's first-strand synthesis primer was used (GAGAGAGAGAGAGAGACTACTGCTGAG(T)-18). After second-strand synthesis is, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size fractionated with a 400bp cutoff, using a SizeSep 400 spun column from Pharmacia. The column eluent was then ligated into Stragene's pBluescript II Xr predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 40 a 60 c 52 g 85 t

ORIGIN

Query Match 64.0%; Score 16; DB 111; Length 237;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttctgcatctgctg 16
|||||

Db 96 GTTCTGATCTGCTGT 111

RESULT 7
LOCUS BE023256 247 bp mRNA EST 21-NOV-2000
DEFINITION sm80a02.y1 Gm-c1015 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1015-6339 5', mRNA sequence.
ACCESSION BE023256

VERSION BE023256.1 GI:8285697
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 247)
 REFERENCE Shoemaker R., Kelm, P., Vodkin, L., Erpelidg, J., Corvelli, V., Khanna A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 TITLE JOURNAL
 COMMENT Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Insert Length: 438 Std Error: 0.00
 Seq primer: -40RP from Gibco.
 FEATURES
 source
 Location/Qualifiers
 1..247
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-6339"
 /clone_1lb="Gm-cl015"
 /tissue_type="Mature flowers, field grown plants"
 /lab_host="X110-Gold"
 /note="Vector: Bluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into X110-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg."
 BASE COUNT 39 a 65 c 57 g 77 t 9 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 162; Length 247;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttgctgagatctgctg 16
 ||||||||||||||||
 Db 128 GTTGCTGATGCTGCTG 143
 RESULT 8
 LOCUS BF14722/c 274 bp mRNA EST 28-NOV-2000
 DEFINITION UI-R-BJ2-bow-c-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
 ACCESSION BF14722
 VERSION BF14722.1 GI:11402711
 KEYWORDS EST.

SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 274)
 REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found. Not a site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No.
 FEATURES
 source
 Location/Qualifiers
 1..274
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ2-bow-c-11-0-UI"
 /clone_1lb="UI-R-BJ2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_SEQ=None found"
 BASE COUNT 54 a 96 c 106 g 18 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 148; Length 274;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttgctgagatctgctg 16
 ||||||||||||||||
 Db 246 GTTGCTGATGCTGCTG 231
 RESULT 9
 LOCUS BB455974 287 bp mRNA EST 21-JUL-2000
 DEFINITION BB455974 RIKEN full-length enriched, 12 days embryo spinal ganglion
 Mus musculus cDNA clone D130043H15 3', mRNA sequence.
 ACCESSION BB455974
 VERSION BB455974.1 GI:9351467
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 287)
 REFERENCE Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Atakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

REFERENCE 1 (bases 1 to 293)
AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ulliu, E., Melville, S., Donelson, J., Fraser, C., and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other-SSS: Sheared DNA-67G1.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through Research Genetics, Alabama, USA. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mbd/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..293
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-67G1"
/note="Vector: pUC18; Site:1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Bartell, Oxford University Press, 1999)."

BASE COUNT
ORIGIN 38 a 73 c 78 g 104 t

Query Match 64.0%; Score 16; DB 240; Length 293;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctgctgttgaagc 24
|||||
DB 149 ATCTGCTGTTGAAGC 164

RESULT 12
AM306120 301 bp mRNA EST 20-JAN-2000
LOCUS
DEFINITION se46b06.y1 Gm-cl017 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
ACCESSION AM306120
VERSION
KEYWORDS
SOURCE EST.
ORGANISM soybean.
glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 301)
Shoemaker, R., Keim, P., Vodkin, L., Erpelidng, J., Corryell, V., Khanna, A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wille, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Riltter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R./Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: 40RP from Gibco.

FEATURES
source
Location/Qualifiers
1..301
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-1932"
/clone_1ib="Gm-cl017"
/issue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II Xr; Site:1: EcoRI; Site:2: XhoI; This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II Xr library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidng."

BASE COUNT
ORIGIN 73 a 48 c 60 g 119 t 1 others

Query Match 64.0%; Score 16; DB 114; Length 301;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactgctg 16
|||||
DB 45 GTTGTGATGATGCTG 60

RESULT 13
BE659333 328 bp mRNA EST 06-SEP-2000
LOCUS
DEFINITION GM700009A20A6 Gm-r1070 glycine max cDNA clone Gm-r1070-3084 3',
ACCESSION BE659333
VERSION
KEYWORDS
SOURCE EST.
ORGANISM soybean.
glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 328)
Erpelidng, J., Raph, C., Shoop, E., Pardinas, J., Liu, L., and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: AM423279 corresponding to Gm-cl015-4654 (5')
Contact: Vodkin, L.O., PI. A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics

University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source

Location/Qualifiers

1..328

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="Gm-r1070-3084"

/clone_lib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, retracted set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were retracted to form library Gm-r1070. The cDNA clones of the retracted Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the Laboratory of Ernest Reitzel, Center for Computational Genomics and Bioinformatics, University of Minnesota. http://www.cbc.umn.edu/research/Projects/soybean/index.html
Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois.
http://www.life.uiuc.edu/bioecch/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 115 a 71 c 76 g 53 t 13 others
ORIGIN

Query Match 64.0%; Score 16; DB 138; Length 328;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagatctgctg 16
|||||
DB 226 gttcctgagatctgctg 211

RESULT 14
AM423279 342 bp mRNA EST 09-FEB-2000
LOCUS sh5d11.y1 Gm-cl015 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl015-4654 5' similar to TR:Q40380 Q40380
ACCESSION ARRBINGALACTAN-PROTEIN PRECURSOR. ; mRNA sequence.
AM423279
VERSION AM423279.1 GI:6951280
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids 1: Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 342)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

TITLE
JOURNAL
COMMENT

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 328.

FEATURES

source

Location/Qualifiers

1..342

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-4654"

/clone_lib="Gm-cl015"

/issue_type="Mature flowers, field grown plants"

/lab_host="Xl10-Gold"

/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into Xl10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 70 a 72 c 70 g 130 t
ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 342;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgagatctgctg 16
|||||
DB 101 gttcctgagatctgctg 116

RESULT 15
C61070/c 360 bp mRNA EST 22-SEP-1997
LOCUS C61070 yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk224c7 5', mRNA sequence.
ACCESSION C61070.1 GI:2419775
VERSION C61070.1
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida; Rhabditioidea ; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
Expression map of the C.elegans genome
Unpublished (1996)
TITLE JOURNAL
COMMENT Contact: Yuji Kohara
Gene Library Lab

National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@nig.ac.jp.

FEATURES

source

Location/Qualifiers
1. 360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK224c7"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 101 a 71 c 76 g 111 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 156; Length 360;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctgcgtttgaagc 24
|||||
Db 26 ATCTGCTGTTGAAGC 11

RESULT 16
BE662000/c 365 bp mRNA EST 08-JAN-2001
DEFINITION bs02d05.y1 Drosophila melanogaster adult testis library Drosophila
melanogaster cDNA clone bs02d05 5', mRNA sequence.
ACCESSION BE662000
VERSION BE662000
KEYWORDS EST.
SOURCE BE662000.1 GI:9992572
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 365)
Andrews, J., Bouffard, G.G., Cheadle, C., Lu, J., Becker, K.G. and
Oliver, B.
Gene discovery using computational and microarray analysis of
transcription in the drosophila melanogaster testis
Genome Res. 10 (12), 2030-2043 (2000)
20568492

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliver@helix.nih.gov,
http://www.nidk.nih.gov/Intram/people/boliver.htm
Tissue isolation and library construction performed at the National
Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
http://www.nidk.nih.gov/Intram/people/boliver.htm). DNA sequencing
and analyses performed by National Institutes of Health Intramural
Sequencing Center (NISC; see http://www.nisc.nih.gov).
Plate: 02 row: d column: 05
Seq primer: M13P1 reverse primer (ABI).
Location/Qualifiers
1. 365

FEATURES

source

/organism="Drosophila melanogaster"
/strain="y1" w[67c1]/y"
/db_xref="taxon:7227"
/clone="bs02d05"
/clone_lib="Drosophila melanogaster adult testis library"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/note="Organ: testis; Vector: pBluescript SK (Stratagene);

Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
day adult y1" w[67c1]/y males raised at 25°C. RNA
isolated using Trizol (Life Technologies) and a single
round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
library constructed using Stratagene ZAP-cDNA synthesis
kit. Oligo dt-primed, size fractionated -1.6 kb, and
directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
Following a single round of amplification pBluescript SK
phagemids were mass excised. A distribution channel for
clones is being sought, but not currently available.
Requests for clones cannot be honored."

BASE COUNT 85 a 129 c 94 g 56 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 138; Length 365;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttgcgtgacgtcg 16
|||||
Db 330 GTTGCGTGACGTGCTG 315

RESULT 17
A1736967 422 bp mRNA EST 06-DEC-1999
LOCUS SB35C02.Y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1013-27 5', similar to TR:Q40380 Q40380 ARABINOGLACTAN-PROTEIN
PRECURSOR. ;, mRNA sequence.
ACCESSION A1736967
VERSION A1736967.1 GI:5058435
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 422)
Shoemaker, R., Keim, P., Vothkin, L., Erpelting, D., Corvelli, V., Khana
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 409.
Location/Qualifiers
1. 422

FEATURES

source

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-27"
/clone_lib="Gm-c1013"
/tissue_type="Whole seedlings, 2-3 week old seedlings,
greenhouse grown"
/lab_host="X110-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI. This cDNA library was constructed from mRNA isolated
from whole seedlings of 2-3 week old greenhouse grown

plants. The cDNA library was prepared using the Stratagene pluscript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluscript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg.

BASE COUNT 83 a 105 c 91 g 138 t 5 others
ORIGIN

Query Match 64.0%; Score 16; DB 24; Length 422;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctgctg 16
|||||
Db 183 GTTGCTGAGTCTGCTG 198

RESULT 18
AM704839 434 bp mRNA EST 21-NOV-2000
LOCUS
DEFINITION AK40B11.Y1 Gm-c1019 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
AM704839 Gm-c1019-5014 5', mRNA sequence.
ACCESSION
VERSION AM704839.1 GI:7589057
KEYWORDS
SOURCE soybean.
ORGANISM
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE
AUTHORS Shoemaker, R., Kelm, P., Vodka, L., Erpelidg, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Mattern, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
1. 434
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-5014"
/clone_id="Gm-c1019"
/tissue_type="Immature seed coats of greenhouse grown plants"
/lab_host="DH10B (Gibco BRL)"
/note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of

greenhouse grown plants. The library was prepared using the Life Technologies pluscript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodka and Dr. Ann Khanna.

BASE COUNT 57 a 169 c 90 g 118 t
ORIGIN

Query Match 64.0%; Score 16; DB 119; Length 434;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctgctg 16
|||||
Db 398 GTTGCTGAGTCTGCTG 413

RESULT 19
A1293622 436 bp mRNA EST 01-DEC-1998
LOCUS
DEFINITION LP06802.5prime LP Drosophila melanogaster larval-early pupal POT2
A1293622 Drosophila melanogaster cDNA clone LP06802 5prime, mRNA sequence.
ACCESSION
VERSION A1293622.1 GI:3943029
KEYWORDS
SOURCE fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
Unpublished (1997)
BDF/HMT Drosophila EST Project
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSJ, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 68 row: A column: 2
High quality sequence stop: 390.

TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
1. 436
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP06802"
/clone_id="LP Drosophila melanogaster larval-early pupal POT2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: POT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library.

BASE COUNT 111 a 122 c 111 g 92 t
ORIGIN

Query Match 64.0%; Score 16; DB 18; Length 436;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctgctg 16

Db 328 GTTGCTGATCTGCTG 313

RESULT 20
LOCUS BE190258

DEFINITION BE190258 441 bp mRNA EST 22-JUN-2000
Gm-c1035-y1 Gm-c1035 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

ACCESSION BE190258
VERSION BE190258.1 GI:8669151

KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 441)
Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marita, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 440.

FEATURES

SOURCE

Location/Qualifiers
1..441
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1035-2848"
/clone_1ib="Gm-c1035"
/tissue_type="Immature leaves of greenhouse grown plants"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from immature leaves (unfurled trifoliolate) of greenhouse grown plants that were 2 weeks old. The library was prepared using the Life Technologies Superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. SalI linker adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT 102 a 63 c 85 g 190 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 164; Length 441;
Best Local Similarity 100.0%; Pred. No. 48;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactgctg 16
|||||

Db 48 GTTGCTGATCTGCTG 63

RESULT 21
LOCUS AL373871

DEFINITION AL373871 470 bp mRNA EST 03-AUG-2000
MTRB03D01F1 MTRB Medicago truncatula cDNA clone MTRB03D01 T3, mRNA

ACCESSION AL373871
VERSION AL373871.1 GI:9673623

KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 470)
Journet, E.P., Crespeau, H., Van-Tuinen, D., Gouzy, J., Jallou, O., Nlebel, A., Carreau, V., Chatagnier, O., Kahn, D., Gianinazzi-Pearson, V. and Gamas, P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
Unpublished (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr
Contact: Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Molculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email: Mt-est@toulouse.inra.fr Website: http://sequence.toulouse.inra.fr/Mtruncatula.html).
Location/Qualifiers
1..470
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MTRB03D01"
/clone_1ib="MTRB"
/tissue_type="Symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from extractase and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exsacit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

FEATURES

SOURCE

BASE COUNT 85 a 137 c 68 g 180 t
ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 470;
Best Local Similarity 100.0%; Pred. No. 48;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactgctg 16
|||||

RESULT 22

LOCUS AM704996

DEFINITION AM704996 474 bp mRNA EST 18-APR-2000
SK41c02.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

ACCESSION GM-c1019-5115 5', mRNA sequence.
 VERSION AM704996
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max

REFERENCE
 AUTHORS Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 474)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Corryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

FEATURES
 source
 Location/Qualifiers
 1..474
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: GM-c1019-5115"
 /clone_1lb="GM-c1019"
 /tissue_type="Immature seed coats of greenhouse grown plants"
 /lab_host="DH10B (Gibco BRL)"
 /note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Ila Vodkin and Dr. Anu Khanna."

BASE COUNT
 ORIGIN
 66 a 175 c 104 g 129 t

Query Match 64.0%; Score 16; DB 119; Length 474;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtctgcatctgctg 16
 |||
 DB 398 GTTCTGATCTGCTG 113

RESULT 23
 LOCUS AL379973 481 bp mRNA EST 03-AUG-2000
 DEFINITION MCB848E08F1 MCB8 Medicago truncatula cDNA clone MCB848E08 T3, mRNA sequence.

ACCESSION AL379973
 VERSION AL379973.1 GI:9679725
 KEYWORDS barrel medic.
 SOURCE Medicago truncatula
 ORGANISM Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 481)
 Journeel, E.P., Crespeau, H., van Tuinen, D., Gouzy, J., Jallou, O., Niebel, A., Carreau, V., Chataigner, O., Kahn, D., Gianinazzi-Pearson, V. and Gamas, P.

TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
 JOURNAL Unpublished (2000)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Contact: Pascal Gamas and Etienne-Pascal Journeel, Laboratoire de Biologie Molculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email: mt-est@toulouse.inra.fr Website: http://sequence.toulouse.inra.fr/Mtruncatula.html).
 Location/Qualifiers
 1..481
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MCB848E08"
 /clone_1lb="MCBB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
 /note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into uni-zap XR vector from Stratagene and packaged using GigaPack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT
 ORIGIN
 89 a 152 c 64 g 176 t

Query Match 64.0%; Score 16; DB 105; Length 481;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtctgcatctgctg 16
 |||
 DB 385 GTTCTGATCTGCTG 400

RESULT 24
 LOCUS BF461669/c 497 bp mRNA EST 04-DEC-2000
 DEFINITION UT-M-CGDP-bmu-g-12-0-UI-s1 NRH BMAP Ret4.S2 Mus musculus cDNA clone
 UT-M-CGDP-bmu-g-12-0-UI-3', mRNA sequence.
 ACCESSION BF461669.1 GI:11530852
 VERSION BF461669
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

REFERENCE 1 (bases 1 to 497)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chih, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestrail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENERICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers

1..497
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U1-M-CG0p-Dmu-9-12-0-UT"
 /clone_id="NIH_BMAP_Ret4_S2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT713D-Pac (Pharmacia) with a modified polylinker. Site 1: Not 1; Site 2: Eco RI; The NIH BMAP Ret4_S2 library is a subtracted library ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
 TAG_SEQ=None found"
 BASE COUNT 112 a 168 c 171 g 44 t 2 others
 ORIGIN

Query Match 64.0%; Score 16; DB 149; Length 497;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgctgctg 16
 |||
 Db 246 GTTCTGATCTGCTG 231

RESULT 25
 BE515828 502 bp mRNA EST 08-AUG-2000
 LOCUS WHE0606.B07.C14ZA Wheat ABA-treated embryo cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE0606_B07_C14, mRNA sequence.
 ACCESSION BE515828
 VERSION BE515828.1 GI:9739842
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticaceae; Triticum.
 1 (bases 1 to 502)
 REFERENCE Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey,S.D. and Walker-Simmons,M.K.
 TITLE The structure and function of the expressed portion of the wheat genomes - ABA-treated embryo library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Glin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
 Seq primer: Clontech Matchmaker 3' Ad primer.
 Location/Qualifiers

FEATURES
 source

1..502
 /organism="Triticum aestivum"
 /cultivar="Brevor (soft, white, winter, common wheat)"
 /db_xref="taxon:4565"
 /clone="WHE0606.B07_C14"
 /clone_id="Wheat ABA-treated embryo cDNA library"
 /tissue_type="Seed embryo"
 /dev_stage="Mature dormant seeds"
 /lab_host="E. coli DH12S"
 /note="Vector: pGAD10; Site 1: EcoRI; Site 2: XhoI; Embryos were cut from mature, dormant seeds and imbedded in 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7, for 12 hr at 22 C. The tissue, total RNA, and poly(A) RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA ARS, Washington State Univ., Pullman, Washington 99164-6420. A cDNA library was made by Clontech using a combination of random and oligo dT primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
 BASE COUNT 103 a 108 c 128 g 162 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 136; Length 502;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctgctgttgaagc 24
 |||
 Db 189 ATCTGCTGTGTAAGC 204

RESULT 26
 AL383090 503 bp mRNA EST 03-AUG-2000
 LOCUS MTCB11G02R1 MTCB Medicago truncatula cDNA clone MTCB11G02 T7, mRNA
 DEFINITION sequence.
 ACCESSION AL383090
 VERSION AL383090.1 GI:9682841
 KEYWORDS EST.
 SOURCE batrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 503)
 REFERENCE Journet,E.P., Crespeau,H., van Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
 TITLE Medicago truncatula ESTs from endomycorrhizal roots
 JOURNAL Unpublished (2000)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: secrete@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : mc-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
 Location/Qualifiers
 1..503
 source

/organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MTBC11G02"
 /clone_lib="MTBC"
 /tissue_type="arbuscular mycorrhizal"
 /dev_stage="harvested 3 weeks post inoculation with Glomus intraradices"
 /note="Vector: pBluescript pSK; Site-1: EcoRI; Site-2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epsilones soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate IPR8). The plants were watered every day and twice a week with a modified nutrient long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exasit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin."

BASE COUNT 107 a 139 c 76 g 181 t
 ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 503;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctgagatcgtcgt 16
 |||||||
 Db 237 GTTGCTGATCTGCTG 252

RESULT 27
 AL373872 517 bp mRNA EST 03-AUG-2000
 LOCUS MEBB03D01.R1 MTBB Medicago truncatula cDNA clone MEBB03D01 T7, mRNA
 DEFINITION sequence.
 ACCESSION AL373872
 VERSION AL373872.1 GI:9673624
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 517)
 REFERENCE Journal,E.P., Crespeau,H., van Tuinen,D., Gouzy,D., Jalllon,O., Nibbel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianlinazzi-Pearson V. and Gamas,P.
 TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
 JOURNAL unpublished (2000)
 COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
 FEATURES
 Source 1..517
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"

/clone="MTBB03D01"
 /clone_lib="MTBB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
 /note="Vector: pBluescript pSK; Site-1: EcoRI; Site-2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exasit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT 104 a 147 c 71 g 195 t
 ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 517;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctgagatcgtcgt 16
 |||||||
 Db 274 GTTGCTGATCTGCTG 289

RESULT 28
 A0692054 517 bp DNA GSS 06-JUL-1999
 LOCUS HS_5408_A2_E07-T7A RPT-11 Human Male BAC library Homo sapiens
 DEFINITION genomic clone Plate-984 Col-14 Row-I, DNA sequence.
 ACCESSION A0692054
 VERSION A0692054.1 GI:5382302
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 517)
 REFERENCE Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Fuliong,T., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPT-11. For BAC library availability, please contact Pliet de Jong (pliet@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 964 Row: I Column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 517.
 FEATURES
 Source 1..517
 Location/Qualifiers
 1..517
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

```

/clone="Plate=984 Col=14 Row=1"
/clone.lib="RRC1-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
BASE COUNT      154 a      96 c      83 g      178 t      6 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 atcgtctgtttgaagc 24
      |||||
Db 411 ATCTGCTGTTTGAGGC 426

RESULT 29
BF419551/c
LOCUS      BF419551      521 bp      mRNA      EST      28-NOV-2000
DEFINITION      UI-R-CAO-bpd-c-11-0-UI-s1 UI-R-CAO Rattus norvegicus cDNA clone
ACCESSION      BF419551
VERSION      BF419551.1 GI:11407540
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 521)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704447
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA-No.

FEATURES
source
Location/Qualifiers
1..521
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAO-bpd-c-11-0-UI"
/clone.lib="UI-R-CAO"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia)
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAO
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
, midbrain, cerebral cortex, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG-Seq=None found"

```

ORIGIN					
Query Match	64.0%;	Score 16;	DB 148;	Length 521;	
Best Local Similarity	100.0%;	Pred. No. 49;			
Matches	16;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	gtgtcctgcatcgcgcg	16		
Db	246	gtttgctgcatcgtcgtg	231		
RESULT 30					
LOCUS	AI854720/c				
DEFINITION	AI854720	522 bp	mRNA	EST	15-JUN-1999
ACCESSION	UI-M-BH0-akc-f-12-0-UI-S1		NIH_BMAP_M_S1	Mus musculus	CDNA clone
VERSION	AI854720				
KEYWORDS	AI854720.1	GI:5498626			
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 522)				
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
	Normalization and subtraction: two approaches to facilitate gene				
	discovery				
JOURNAL	Genome Res.	6 (9),	791-806	(1996)	
MEDLINE	97044477				
COMMENT					
	Contact: Chin, H				
	National Institute of Mental Health				
	6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD				
	20892-9643, USA				
	Tel: 301 443 1706				
	Fax: 301 443 9890				
	Email: MEST@mail.nih.gov				
	Oligo-ct track not found, Not I site shown in beginning of sequence				
	is likely internal to the message. cDNA library preparation: M.B.				
	Soares lab clone distribution: NIH BMAP cDNA clones will be made				
	available by the means that is soon to be determined. When NIH				
	determines the means for distribution of the BMAP cDNA clones, this				
	record will be updated accordingly when that means is determined.				
	Seq primer: M13 forward				
	POLYA-No.				
FEATURES					
source	Location/Qualifiers				
	1..522				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db.xref="taxon:10090"				
	/clone="UI-M-BH0-akc-f-12-0-UI"				
	/clone_11p="NIH_BMAP_M_S1"				
	/dev_stage="27-32 days"				
	/lab_host="DH10B (Life Technologies)"				
	/note="Vector: pT713D-Pac (Pharmacia) with a modified				
	polylinker; Site 1: Not I; Site 2: Eco RI; The				
	NIH_BMAP_M_S1 library is a subtracted library derived from				
	a mixture of normalized libraries from ten regions of the				
	mouse brain (cerebellum, brain stems, olfactory bulbs,				
	hypothalamus, cortex, amygdala, basal ganglia, pineal				
	gland, striatum, hippocampus). The driver used for				
	subtraction consisted of a pool of 20,000 cDNA clones				
	obtained from non-normalized and normalized libraries of				
	these ten regions of the mouse brain.				
	TAG_LIB=NIH_BMAP_M_S1				
	TAG_TISSUE=hypothalamus				
	TAG_SEQ=CGGTA"				
BASE COUNT	118 a	174 c	179 g	51 t	
ORIGIN					
Query Match	64.0%;	Score 16;	DB 102;	Length 522;	
Best Local Similarity	100.0%;	Pred. No. 49;			

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgagtcgtcgt 16
 |||
 Db 245 GTTGCTGATCTGCTG 230

RESULT 31
 AL379974 525 bp mRNA EST 03-AUG-2000
 LOCUS MTBBA8E08R1 Medicago truncatula cDNA clone MTBBA8E08 T7, mRNA
 DEFINITION sequence.
 ACCESSION AL379974
 VERSION AL379974.1 GI:9679726
 KEYWORDS EST.
 SOURCE bareil medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 525)
 REFERENCE 1
 JOURNAL Journef,E.P., Crespeau,H., Van-Tuinen,D., Gouzy,J., Jallion,O.,
 AUTHORS Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson
 V. and Gamas P.
 Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
 nodules
 TITLE Unpublished (2000)
 JOURNAL Contact: Genoscope
 COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journef, Laboratoire de
 Biologie Moléculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 M-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).
 FEATURES
 source
 1. 525
 Location/Qualifiers
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone_id="MTBBA8E08"
 /clone_1lb="MTBB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with
 Sinorhizobium meliloti"
 /note="Vector: pBluescript psk; Site_1: EcoRI; Site_2:
 XhoI; Plants were grown in an aeroponic chamber on
 nitrogen-rich medium for 21 days. Three days before
 inoculation with Sinorhizobium meliloti, the medium was
 replaced by N-free medium. Root nodules (+ short adjacent
 root segments) were harvested 4 days post inoculation.
 cDNA was prepared from polyA+ enriched RNA. The cDNA was
 directionally ligated into uni-zip XR vector from
 Stratagene and packaged using Gigapack Gold packaging
 extracts. Plasmids containing cDNA inserts were
 mass-excised from phage stocks using Exsist helper phage
 and propagated in SOLR cells. Clone ordering and
 sequencing was performed by the Centre National de
 Sequencage (Genoscope, Evry, France)."
 BASE COUNT 108 a 148 c 81 g 188 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 105; Length 525;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttcgtgagtcgtcgt 16
 |||
 Db 255 GTTGCTGATCTGCTG 270

RESULT 32
 AM395303
 LOCUS
 DEFINITION
 AM395303 531 bp mRNA EST 17-JUL-2000
 s146d09.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-c1017-4746 5' similar to SW:R8B9.DROME P36558 DNA-DIRECTED RNA
 POLYMERASE II 15.1 KD POLYPEPTIDE ;, mRNA sequence.
 ACCESSION AM395303
 VERSION AM395303
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 531)
 REFERENCE 1
 JOURNAL Shoemaker,R., Kelm,P., Vodkin,L., Erpelnding,J., Coryell,V., Khanna
 AUTHORS A., Bolla,B., Maria,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 JOURNAL Contact: Shoemaker R/Public Soybean EST Project
 COMMENT Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
 427-3334 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert length: 743 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 342.
 FEATURES
 source
 1. 531
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone_id="GENOME SYSTEMS CLONE ID: Gm-c1017-4746"
 /clone_1lb="Gm-c1017"
 /tissue_type="Vegetable buds of field grown plants"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from vegetable buds of field grown plants. The cDNA
 library was prepared using the Stratagene pBluescript II
 XR library construction kit. Complementary DNA was
 synthesized from mRNA using a primer consisting of a poly
 (dT) sequence with a XhoI restriction site. EcoRI adapters
 were ligated to the blunt-ended cDNA fragments followed by
 XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into XL10-Gold host cells. This library was
 constructed by Dr. Randy Shoemaker and Dr. John
 Erpelnding."
 BASE COUNT 159 a 96 c 107 g 169 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 115; Length 531;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttcgtgagtcgtcgt 16
 |||
 Db 60 GTTGCTGATCTGCTG 75

RESULT 33
 A0889100/c
 LOCUS
 DEFINITION HS.2204_A1-C07_MR C1T Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
 ACCESSION A0889100
 VERSION A0889100.1 GI:6345290
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 575)
 Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahaitas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2204 row: E column: 13
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 555.
 FEATURES
 source
 Location/Qualifiers
 1..555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate=2204 Col=13 Row=E"
 /sex="Male"
 /note="Organ: Sperm; Vector: pBelorAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 131 a 135 c 147 g 124 t 18 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 235; Length 555;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 10 tctctgttgaagc 25
 ||||||||||||
 Db 497 TCTGCTCTTGACGC 482

RESULT 34
 A0889100
 LOCUS
 DEFINITION HS.2204_A1-C07_MR C1T Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
 ACCESSION A0889100
 VERSION A0889100.1 GI:6345290
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 575)
 Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahaitas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2204 row: E column: 13
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 555.
 FEATURES
 source
 Location/Qualifiers
 1..555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate=2204 Col=13 Row=E"
 /sex="Male"
 /note="Organ: Sperm; Vector: pBelorAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 131 a 135 c 147 g 124 t 18 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 235; Length 555;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 10 tctctgttgaagc 25
 ||||||||||||
 Db 497 TCTGCTCTTGACGC 482

JOURNAL
 COMMENT truncated
 Unpublished (2000)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 Minnesota EST name: M252705e
 TIGR sequence name: MTCAP957K
 More information is available at:
 'http://chrysie.tamu.edu/medicago/'
 Seq primer: SKmod (CTA gAA gAT CC).
 FEATURES
 source
 Location/Qualifiers
 1..574
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pgVN-48022"
 /clone_lib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the uni-ZAP Xk vector from stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells."
 BASE COUNT 110 a 165 c 80 g 218 t 1 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 118; Length 574;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 gttgtcgatctgctg 16
 ||||||||||||
 Db 370 GTTGTGATCTGCTG 385

RESULT 35
 BE999070
 LOCUS
 DEFINITION BE999070 GVSN Medicago truncatula cDNA clone pgVSN-15C19, mRNA sequence.
 ACCESSION BE999070
 VERSION BE999070.1 GI:10699346
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 575)
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
 ESTs from senescent nodules of Medicago truncatula
 Unpublished (2000)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M273048e TIGR sequence name:

MTKAY22K More information is available at:
<http://chryslie.tamu.edu/medicago>
 Seq primer: Skmod (CTA gAA CTA gTg gAT CC).
 Location/Qualifiers

FEATURES
 source
 1.575
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone_lib="pGVSN-15C19"
 /clone_1ib="GVSN"
 /issue_type="senescent root nodules"
 /dev_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."
 BASE COUNT
 103 a 164 c 82 g 226 t
 ORIGIN

Query Match 64.0%; Score 16; DB 143; Length 575;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgtgagatcgtctg 16
 |||||
 Db 347 GTTGCTGATCTGCTG 362

RESULT 36
 BE525150 580 bp mRNA EST 11-DEC-2000
 LOCUS
 DEFINITION
 BE525150-1 GI:11633117
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 580)
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES
 source
 1.580
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 Email: mscoares@blue.weeg.iowa.edu
 cDNA library preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LBNL (lbnlimage.lbnl.gov). IMAGE ID= 1795307
 Seq primer: M13 Forward.
 Location/Qualifiers
 1.580
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"

/db_xref="taxon:10116"
 /clone="UI-R-AF0-yd-e-05-0-UI"
 /clone_1ib="UI-R-AF0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF0
 library is a non-normalized library constructed from 15
 dpc rat atriocentricular (AV) canal. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bernaldo, Lennon and Soares, Genome Research 6: 791-806
 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa."
 BASE COUNT
 155 a 146 c 147 g 132 t
 ORIGIN

Query Match 64.0%; Score 16; DB 150; Length 580;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtgagatcgtctgt 18
 |||||
 Db 321 TCCTGATCTGCTGTGT 336

RESULT 37
 AW734894 591 bp mRNA EST 24-APR-2000
 LOCUS
 DEFINITION
 AW734894.1 GI:7640531
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 380.

FEATURES
 source

1.591
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-10014"
 /clone_1ib="Gm-c1016"
 /issue_type="Immature flowers of field grown plants"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated

from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 121 a 180 c 103 g 185 t 2 others

Query Match 64.0%; Score 16; DB 120; Length 591;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtctgattctgtctg 16
|||||

Db 330 GTTGCTGATCTGCTG 345

RESULT 38
A2267971 601 bp DNA GSS 26-JUL-2000
LOCUS RPT-23-103L11.TJ RPT-23 Mus musculus genomic clone RPT-23-103L11
DEFINITION , DNA sequence.
ACCESSION A2267971
VERSION A2267971.1 GI:9481584
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 601)
AUTHORS Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Aklnret, B., Levins, M., McGam, S., Tsegaye, G., Geer, K., Krol, W., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPT-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPT-23-103L11.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPT-23. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resaca ch Genetics (<http://resgen.com>). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 103 row: L column: 11
Seq primer: SP6
Class: BAC ends.

FEATURES

Source Location/Qualifiers
1..601
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPT-23-103L11"
/clone_lib="RPT-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 189 a 92 c 111 g 209 t

Query Match 64.0%; Score 16; DB 240; Length 601;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtctgattctgtctg 17
|||||

Db 354 TTGCTGATCTGCTG 369

RESULT 39
AM267738 609 bp mRNA EST 07-SEP-2000
LOCUS EST305866 DSIR Medicago truncatula cDNA clone pDSIR-717, mRNA
DEFINITION sequence.
ACCESSION AM267738
VERSION AM267738.1 GI:6654694
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago. 1 (bases 1 to 609)
AUTHORS Fedorova, M., Pierson, B.L., Sanao, D.A., Vance, C.P., Gantt, G.S., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, L.E. and Fraser, C.M.
TITLE ESTs from roots of Medicago truncatula after inoculation with Phytophthora medicaginis
JOURNAL Unpublished (1999)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
TITLE TIGR sequence name: MTBA527K
COMMENT More information, including clone ordering, is available at: <http://chrysis.tamu.edu/medicago>
Seq primer: SKmd (CTA GAA CTA gtc gat CC).

FEATURES

Source Location/Qualifiers
1..609
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pDSIR-717"
/clone_lib="DSIR"
/tissue_type="roots infected with Phytophthora medicaginis"
/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOR cells. Note: EST may be of fungal origin."

BASE COUNT 117 a 172 c 82 g 238 t

Query Match 64.0%; Score 16; DB 113; Length 609;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgactctgctg 16
 |||||
 Db 381 GTTGCTGATCTGCTG 396

RESULT 40
 AM980891 616 bp mRNA EST 07-SEP-2000
 LOCUS EST392044 GVN Medicago truncatula cDNA clone pGVN-60C11, mRNA
 DEFINITION
 sequence.
 ACCESSION AM980891 GI:8172436
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 616)
 REFERENCE
 Pedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gant, J.S., Peng
 H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
 Holt, I.E. and Fraser, C.M.
 ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula
 TITLE
 Unpublished (2000)
 JOURNAL
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M256725e
 TIGR sequence name: MTCBM18TK
 More information is available at: <http://chryslie.tamu.edu/medicago>
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
 FEATURES
 source
 1. 616
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pGVN-60C11"
 /clone_1ib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month
 post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 effective root nodules harvested one month post
 inoculation with Sinorhizobium meliloti. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells."

BASE COUNT 118 a 173 c 82 g 243 t
 ORIGIN

Query Match 64.0%; Score 16; DB 123; Length 616;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgactctgctg 16
 |||||
 Db 388 GTTGCTGATCTGCTG 403

RESULT 41
 BG585328 618 bp mRNA EST 11-APR-2001
 LOCUS EST487092 MHAM Medicago truncatula/Glommus versiforme mixed EST
 DEFINITION
 library cDNA clone PMHAM-22F21 5' end, mRNA sequence.
 ACCESSION BG585328
 VERSION BG585328.1 GI:13600392
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago truncatula/Glommus versiforme mixed EST library.
 Medicago truncatula/Glommus versiforme mixed EST library
 Eukaryota; mixed EST libraries.
 1 (bases 1 to 618)
 REFERENCE
 Harrison, M.J., Liu, J., Town, C.D., Van Aken, S., Uutterback, T., Cho, J.
 and Fraser, C.M.
 ESTs from roots of Medicago truncatula after colonization with
 Glommus versiforme, 2001
 TITLE
 Unpublished (2001)
 JOURNAL
 Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Noble EST name: N380802e TIGR sequence name: MNDCC35TK More
 information is available at: <http://www.medicago.org>
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
 FEATURES
 source
 1. 618
 /organism="Medicago truncatula/Glommus versiforme mixed EST
 library"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="PMHAM-22F21"
 /clone_1ib="MHAM"
 /tissue_type="roots colonized with Glommus versiforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glommus versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glommus versiforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."

BASE COUNT 119 a 174 c 84 g 241 t
 ORIGIN

Query Match 64.0%; Score 16; DB 155; Length 618;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgactctgctg 16
 |||||
 Db 390 GTTGCTGATCTGCTG 405

RESULT 42
 AM422111 621 bp mRNA EST 09-FEB-2000
 LOCUS f157a12 y1 Suano Kawakami zebrafish DRB danio rerio cDNA clone
 DEFINITION (HUMAN);, mRNA sequence.
 ACCESSION AM422111 GI:6950043
 VERSION
 KEYWORDS
 SOURCE
 zebrafish.

ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 621)

REFERENCE
Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Rittler,
E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.
Masnu Zebrafish EST Project 1999
Unpublished (1999)

TITLE
JOURNAL
COMMENT
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: 73 bp from Amersham
High quality sequence stop: 504.
Location/Qualifiers
1..621

FEATURES
source
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2641726"
/clone_lib="Sugano Kawakami zebrafish DNA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG);
Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGGGCGCTTTTCTTTTCTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TTGGGCGCTGTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' and primer CTTCTGCTCTAAAGCTCG and 3' end
primer CGACCTGCAGCTGCAGACA. "

BASE COUNT
185 a 174 c 143 g 119 t

ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 621;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtgtgactgtgtgt 17
|||||

Db 138 TTGCTGATCTGCTGT 123

RESULT 43
BE593142 627 bp mRNA EST 18-AUG-2000
LOCUS WSI_98_A09.g1_A002 Water-stressed 1 (WSI) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE593142
VERSION BE593142.1 GI:9848215
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 627)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt

TITLE
JOURNAL
COMMENT
An EST database from Sorghum: water-stressed plants
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 55
High quality sequence stop: 627
POLYA-No.

FEATURES
source
location/Qualifiers
1..627
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WSI)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: lambda Zap; Site 1: XhoI;
Site 2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT
169 a 152 c 124 g 182 t

ORIGIN

Query Match 64.0%; Score 16; DB 137; Length 627;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttgtgtgactgtgtgt 18
|||||

Db 208 TGCTGATCTGCTGT 193

RESULT 44
AZ449344 632 bp DNA GSS 04-OCT-2000
LOCUS AZ449344 Mouse 10kb plasmid UNGC1M library Mus musculus genomic
DEFINITION clone UNGC1M0247J04 R, DNA sequence.
ACCESSION AZ449344
VERSION AZ449344.1 GI:10603038
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 632)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0247 row: j column: 04
Seq primer: CACACGGAACACGATGACG
Class: plasmid ends
High quality sequence stop: 632.

FEATURES

source

Location/Qualifiers

1. 632

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U08C1024704"

/clone_lib="Mouse 10kb plasmid U08C1M library"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g11473211419b/AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT

185 a 167 c 110 g 170 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 244; Length 632;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggaatcgtcgtttaa 22

Db 552 GGATCTGCTGTTCGA 567

RESULT 45

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 640)

AUTHORS

REFERENCE

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 LSA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: http://www.fruitfly.org/EST, estfruitfly.berkeley.edu

Plate: 80 row: A column: 3

High quality sequence stop: 412.

location/Qualifiers

1. 640

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="GH08003"

/clone_lib="GH Drosophila melanogaster head port2"

/sex="male and female"

/dev_stage="adult"

/lab_host="DH5 - alpha"

/note="Organ: head; Vector: port2; Site1: EcoRI; Site2: XhoI; Sized fractionated cDNAs were directly ligated into port2 plasmid cDNA library."

BASE COUNT 160 a 211 c 175 g 94 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 16; Length 640;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggaatctg 16

Db 352 GTTGCTGATCTGCTG 337

Search completed: October 9, 2001, 15:15:26

Job time: 13648 sec

Wed Oct 10 07:45:47 2001

us-09-396-196f-3.011.rst

Page 24

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:44:56 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25
Sequence: 1 ggatctgctgttgagcgcaagcag 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 237205

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_v1: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v1: *
59: gb_v12: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_pr1: *
95: gb_pr2: *
96: gb_pr4: *
97: gb_pr10: *
98: em_ba3: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 B10B gene O
4	25	100.0	5526	2	AF250776	AF250776 Unculture
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

```

9      25      5872      9      A33679
10     25      5872      9      AR101809
11     25      5872      9      AR101810
12     25      5872      9      AR101810
13     25      5872      9      AR101810
14     25      5872      9      AR101810
15     25      5872      9      AR101810
16     25      5872      9      AR101810
17     25      5872      9      AR101810
18     25      5872      9      AR101810
19     25      5872      9      AR101810
20     25      5872      9      AR101810
21     25      5872      9      AR101810
22     25      5872      9      AR101810
23     25      5872      9      AR101810
24     25      5872      9      AR101810
25     25      5872      9      AR101810
26     25      5872      9      AR101810
27     25      5872      9      AR101810
28     25      5872      9      AR101810
29     25      5872      9      AR101810
30     25      5872      9      AR101810
31     25      5872      9      AR101810
32     25      5872      9      AR101810
33     25      5872      9      AR101810
34     25      5872      9      AR101810
35     25      5872      9      AR101810
36     25      5872      9      AR101810
37     25      5872      9      AR101810
38     25      5872      9      AR101810
39     25      5872      9      AR101810
40     25      5872      9      AR101810
41     25      5872      9      AR101810
42     25      5872      9      AR101810
43     25      5872      9      AR101810
44     25      5872      9      AR101810
45     25      5872      9      AR101810

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton D. Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ggaatcgtgttgaagcagcagcag 25
    |||||||||||||||||||||
Db 63 GGATCTGCTGTTGAAGCCAGCAG 87

```

RESULT 2

```

AR034916 AR034916 1041 bp DNA PAT 29-SEP-1999
LOCUS AR034916
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ggaatcgtgttgaagcagcagcag 25
    |||||||||||||||||||||
Db 63 GGATCTGCTGTTGAAGCCAGCAG 87

```

```

RESULT 3
LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
DEFINITION Biot gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
Source Location/Qualifiers
1. 1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24. 1064
/gene="BIOB"
24. 1064
/gene="BIOB"
24. 1064
/gene="BIOB"
/codon_start=1
/transl_table=1
/protein_id="CA00965.1"
/db_xref="GI:490219"
/translation="MAHRRKWTLSQVTELFKPLDLIFEAQOVHRQHPDPROVYST
LSTRTGACPEDEKCYCPQSSRYRTGLAEERLMEVEQVLESARAKAGSTRFCMGAW
KNPERMDPYLEQWQVQKAMGLACWTLCTLSQVQARLANGLDYNNHNDTSPF
YGNITTRVQERLDLEKVRDAGIKVCSGIVGLGETVDRAGLLIOLANLPTPES
VPINMLKVRGTPLANDVDVDFRTTIVARIMPTSVRLSAGREOMNEQTOAMC
FMAGANSIFYGCKLITTPNEDEKDLQLPFRKGLNQQVAVLAGNDEQQOREQALMT
PDDDEVNNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 9; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ggaatcgtgttgaagcagcagcag 25
    |||||||||||||||||||||

```

Db	86	GGATCTGCTGTTTGAAGCGCAGAG	110
RESULT	4		
LOCUS	AF250776		
DEFINITION	AF250776 5526 bp DNA BCT 31-JAN-2001 Uncultured bacterium pCOSH2 hypothetical 17.1 kDa protein in modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA), biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and biotin biosynthesis protein BioC (bioC) genes, complete cds; and dehydrobiotin synthetase BioD (bioD) gene, partial cds.		
ACCESSION	AF250776		
VERSION	AF250776.1	GI:12620124	
KEYWORDS			
SOURCE	uncultured bacterium pCOSH2. Uncultured bacterium pCOSH2 Bacteria; environmental samples.		
ORGANISM			
REFERENCE	1 (bases 1 to 5526) Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streilt,W.R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia Appl. Environ. Microbiol. 67 (1), 89-99 (2001)		
AUTHORS	11133432 2 (bases 1 to 5526) Entcheva,P., Liebl,W. and Streilt,W.R. Direct Submission Submitted (31-Mar-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Griesbachstr. 8, Goettingen 37077, Germany		
TITLE	Location/Qualifiers		
JOURNAL	1. .5526 /organism="uncultured bacterium pCOSH2". /db_xref="taxon:143797" /clone="pCOSH2" /note="unknown organism, cosmid clone derived from environmental consortium" complement(52. .528) /note="ORF1" /codon_start=1 /transl_table=1 /product="hypothetical 17.1 kDa protein in modC-bioA intergenic region" /protein_id="AAG60577.1" /db_xref="gi:12620125" /translation="MKLISNDLRGDKLPHKRVENGNGYDGDNISPLIANDDVPAQTKSFVVCYDPDAPGTSGSMWVWVNLPAOTRYLPQGFSGGLVAMPDGVLTQRTDFGKAGYDGAAPKRGTHRYITFYHALDIDRIDVDSCASGAMVGFNVHFSLASATIMFS" complement(587. .1876) /gene="bioA" /gene="bioA" complement(587. .1876) /gene="bioA" /note="7,8-diaminopelargonic acid synthetase-aminotransferase" synthetase-aminotransferase" /codon_start=1 /transl_table=1 /product="DAPA-aminotransferase BioA" /protein_id="AAG60578.1" /db_xref="gi:12620126" /translation="MTDDLAFDORHIMHPTSMISPLPYVYVYSAEGCELIISDGRRLYDGRSSMMAALHGINPHQLAANKHSQIDAMSHMFEGITTHAPALIELCRKIVANTPQRLKCVFLADSGVAEEVAKRMALQTMQAGEARQRTLTFRNGYHDDDTGAMSVCPDNDSMHSLMKGYLPENLFAAPQASRMDEQWEDRWGVARLMAARHETIAVLTETIVOGAGAGMIRWHPMLRKIRKICDREGILILADELAFQGTOKLFCCEHAETAPDILCLGKALTGQMTLTATLTTRVAVETIENGEGCFMHGPTMGPNLACAANASIALIESGDMOOCDOVADIEVQLREGLAPARDAENVADVRYGALGVVETTPVMMALOKFFVYBQWYIRIRPFGKLTLYMPYITLPQQLQRLTAANRAVDETFECQ" 1963. .3003 /gene="bioB" 1963. .3003 /gene="bioB" /codon_start=1 /transl_table=1 /product="biotin synthase BioB"		
FEATURES			
source			
CDS			
gene			
CDS			
gene			
CDS			

gene		/protein_id="AA660579.1"	/db_xref="GI:12620127"
CDS		/transl_table=11 /product="KAPA synthetase BioF" /protein_id="AA660580.1" /db_xref="GI:12620128"	
		/translation="MSMOEKINNALDARRAADALRRRYPAVGAGMLVADROYLNFSSNDYIGLSHHPOIIRAMOGAOFEGICSGSGSHVGYSVHQALEELAEWIGYSRALPFISEFAANQAVIAAMMAKEDRIADRUSHALESASISPSQLRFRANDYTHLARSLAPCGOGLLVTEGVFSMGSDSAPLAELIQVTQOHNGWLWDADHGTVICEQGKCSLOVKRKPELLVTFEGKFGSGGSAVLCSITVADYLDRHLIYITSMPNQAOLRASIVARREDGEAREKRLVLIHFHPAGVODLPFTLADSCSAIOPLVDNSRAIOLAERLRQGCWVTATIRPTTYPVAGTRALRLTLTAHEMDIDRLLEVLGHCN"	
gene		4141..4896	
CDS		/gene="bioc"	/note="reaction step prior to synthesis of pimeloyl-CoA"
		/codon_start=1	/transl_table=11
		/product="biotin biosynthesis protein Bioc"	/protein_id="AA660581.1"
		/db_xref="GI:12620129"	/translation="MATVNKOATIAAFGRGAALYECHADHROSADVLLAMLPERKYTHVIDAGGPGPMWRHMERRERAQVTAIDLSPMLVYOARKDAADHYLAGDIIESPLATAFTFDLASNLAVOMCGNSTALARIELRYVRKSQVYATFTLVQGSIPELHQMOWDEEPPHANFLRPDEIEOSLNGVHVQHNIQPTLTFMDDLSMRSLKIGATHLHGDPRILLTRSQIORLQAWPQOOGRPYLYHLFLGVIAER"
gene		4883..>5526	
CDS		/gene="biob"	/gene="biob"
		/note="DTB synthetase"	/note="DTB synthetase"
		/codon_start=1	/transl_table=11
		/product="dethiobiocin synthetase Biob"	/protein_id="AA660582.1"
		/db_xref="GI:12620130"	/translation="MLVSKRYETGTDEVKGTVASCALLQAAAGYRTTAGTKPVASGSEKTPEGLRNLDALQRNSSLOLDAYATVPYFAPFTSPHIISAQEGRETSVMSLGNALROADMVLEVGAGWFPLESDTFADWTOEOLPVLILVGVKLGCIGNHAMLTAAQIHAGLTLAGVANVDVTPGGKRAEYITTLRMIPABLLGEIPMLEMENP"
BASE COUNT	1274 a 1507 c 1567 g 1178 t		
ORIGIN			
Query Match	100.0%; Score 25; DB 2; Length 5526;		
Best Local Similarity	100.0%; Pred. No. 0.00011;		
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 ggatctgctgtttgaagcgacagcag 25		
DB	2025 GCATCTGCTGTGTTGAAGCGACAGCAG 2049		
RESULT	5		
ECOBIO	5793 bp DNA	BCT	28-FEB-1994
LOCUS	E.coli 7, 8-diamino-pelargonic acid (biob), biotin synthetase		
DEFINITION	(biob), 7-keto-8-amino-pelargonic acid synthetase (biof), bioc		

ACCESSION J04423
 VERSION 7.8-diamino-pelargonic acid aminotransferase;
 KEYWORDS 7-keto-8-amino-pelargonic acid synthetase; bioa gene; bioa gene;
 SOURCE bioa gene; bioa gene; bioa gene; bioa gene; bioa gene; bioa gene;
 ORGANISM Escherichia coli (strain K-12) DNA.
 REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
 A.Otsuka, 09-NOV-1988.
 FEATURES
 SOURCE
 1. 5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=1
 /protein_id="AA023513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRDGLPHRHVFNCGYDQDNIISPLAMDDVPACTK
 SPVATYCDPDPAPTHRTITFYVHALDRLIDVDGASAMVGFVNHHSLSASITAFMS"
 complement(633..1925)
 /gene="bioa"
 complement(633..1925)
 /gene="bioa"
 complement(633..1925)
 /codon_start=1
 /transl_table=1
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA023514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFDQRIHMPYTSPLPYVPAVSAEGCLILDSGR
 LVDMSSWMAAIGHYVNPOLNAAMKSOIDMSHMGGITTHAPATLCKRIAMSGRN
 ALDCEVLADSGVAEVAEMKALQYWAQASORPLTRNGHGTTCAMSYCDPDN
 SMHSIMKGYLPENLAPAPQSRMDGEMDEDMGRLMAHHEHETAAVILETVGA
 GMAKMHPEWIKRIKICDREGILLIADETATGFRGKLFACEHETAIADILCLGKA
 LTGGWTLSTLTREVAETISNGEAGCEMHGTFPGNPLACAAANASLAIIESGDMO
 QOYADIEVOLIREQALAPARDAEMVADVIVGALGVETTHPVNMAALQKFEVQGVIR
 PFGKILYLMPPYIILPOLORLTAAVNRAVQDEFPCQ"
 2012..3052
 /gene="bioa"
 2012..3052
 /gene="bioa"
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="AA023515.1"
 /db_xref="GI:145425"
 /translation="MAHRRPWLISOYTELEKPLDLLEAOVHRQHPDPPOVST
 LLSIKTGACPECKVCPTSRKTKTGLAEERLEVEVLSARAKANAAGSTRCAAM
 KNHERMDPYLEDMOVGVKAMGLACMTIGTLESQAQALNAGLIDYVNHNDTSPEP
 YGNILITRTYQERLDLEKRVDAIGIVCGSGIVGGEIVKDBAGLILLOLAN.PTPSES
 VPIMLVYKGTPLADNDVDADFRTITAVARIMMPTSYVRLSAGREOMNOCOLAMC
 PMAGANSIFYGKILTTNPPEEDKQLFLRKGLNPOQTAVALAGNEOQRLQALMT
 PPTDEYNAAL"
 3049..4203
 /gene="bioa"
 3049..4203
 /gene="bioa"
 /codon_start=1
 /transl_table=1

/product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA023516.1"
 /db_xref="GI:145426"
 /translation="MSWDEKINAAIDARRADALRRYPVAGAGRWLVADROLYNF
 SSNDYGLSHHPQIIRAMQOGAEOFGIGSGGSHSGSVVQALAEELAEALGYSSRA
 LLEISGFANQAVIAMMAKEDRIADRLSHASLEAASLSPSRRRAHNDVTLAR
 LIASPCGOOMVYTEGVFSMDSDSAPLAEIOOVTOOHNGWMLVDDAHGRTYGEORG
 SCLOKRVPELLVTFPEKGFVSGAAYVCSSTVADYILOFARHLIYXSMRPAQOAL
 RASIAVIRSDGDAAREKALITPRAGVODLPPTLADSCAIOPLIVGDSRALOL
 AKRLQOGCVVNAIRPVPAGTARLRLITAAHEMODIRLEVLHGNG"
 4190..4945
 /gene="bioa"
 4190..4945
 /gene="bioa"
 /codon_start=1
 /transl_table=1
 /protein_id="AA023517.1"
 /db_xref="GI:145427"
 /translation="MATVKNQAIIAAFGRAAAHYBOHADLQOSADALLAMLPORRYT
 HYLDAGCGGMSRHRERHRAQVTLADLSPMLVQAROKDAADHYLAGDISELPATA
 TFDLAMSNLAVOMCGNLSTALRELYRVVRPGVAFVTLVYOGSLPERHOAOVADERP
 HANRFLPDEIRQSILNGVYHOHHIOPIITLMPDALSAMRSKIGATLHGRDRRL
 TRSOLQRLQIAMPQOQGRPLTYHLFLGYIARE"
 4938..5597
 /gene="bioa"
 4938..5597
 /gene="bioa"
 /codon_start=1
 /transl_table=1
 /product="dethiobiotin synthetase"
 /protein_id="AA023518.1"
 /db_xref="GI:145428"
 /translation="MSKRYVYGTDEVEKTVASCALLQAAKRAGYRTAGYKPYVAGS
 EKTPEGLNSDALALORNSLQLDYATVATVPTFAEPTSPHIIISOEGRISLVMSAG
 LRALQOQADWVVEGACGMFTPLSDTFEFAWVVOEOLPVLIVYGKGCINHMALTA
 QVIOHAGLTLGAVANDVTPPGKRRAEYMTLTIRRCRBERBPGLOKIOKMOPES"
 7

BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of hpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
 Best local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgtttgaagcgacagcag 25
 |||||||||||||||||||||||||
 DB 2074 ggatctgctgtttgaagcgacagcag 2098

RESULT 6
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS Sequence 1 from Patent WO9408023.
 DEFINITION A38246
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 ORGANISM
 Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 REFERENCE Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 AUTHORS BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 TITLE PATENT: WO 9408023-A 1 14-APR-1994;
 JOURNAL LONZA AG (CH)
 COMMENT
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331

Other publication JP 8501694F 960227 .

FEATURES

source location/Qualifiers

1..3672

/organism="Escherichia coli"

/strain="DSM498"

/db_xref="taxon:562"

/clone="PB030A-15/9"

1..96

/function="PROMOTER PTAC"

/evidence=experimental

23..28

45..49

/strand.name="PROMOTER PTAC"

/evidence=experimental

105..119

/strand.name="BIOB RBS NO.9"

/evidence=experimental

117..1157

/gene="BIOB"

117..1157

/gene="BIOB"

/codon_start=1

/transl_table=11

/number=1

/evidence=experimental

/product="BIOTIN SYNTHASE"

/protein_id="CAA02325.1"

/db_xref="GI:2294845"

/translation="MAHRPRWTLQVATLFEKPLLDLLEPAQGVHROHPVOYST
LISRTACPEDCKYCPQSSRYTGLAEPLMEVEOYLESARKAKAGSTRPGMGAM
KNPERMPYLEQWVGCVKAMGLEACHTLGLTLESCQRLANGLDYVHNHLDTSPEF
YGNITITRTYQERLDTLEKVDAGIKVCSGIVGLGTVADRAGLLQVLAHPTPEPS
VPIHMLYKVGSTPLANDDDVADDFIPTTIVGARIMPTSYRLSAGDEONEDQVAMC
FMGANSIFYGCKLITTPNPEEDKDLQFLFKRLINPQQTAVLAGDMDOORLEQALMT
PDPEYVNAAL"

2284..2297

/strand.name="BIOC RBS"

2295..3050

/gene="BIOC"

2295..3050

/gene="BIOC"

/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"

/codon_start=1

/transl_table=11

/number=3

/product="PROTEIN"

/protein_id="CAA02326.1"

/db_xref="GI:2294846"

/translation="MATVYKQATIAAFGRAAHYEGHADLOROSADLLAMLPRKYT
HYLDAGCGPGMWSHWRERHQAQVYRLDSPMLVQAQRKADHYLAGDILESLATA
TFEDLAWSLAYQWCGNLSYALRELYRYVRKGVYAFITLLVQGSILPELHQANQANDERP
HANFLPDETEQSLNVEVYOHHTQPLTWDDLSMRSLSKIGATHLHGDRPRLT
TRSQLORQLAMPOQGRYPLTYHFLGLVIARE"

3742..3752

/strand.name="BIOA RBS"

3750..5039

/gene="BIOA"

3750..5039

/gene="BIOA"

/strand.name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."

/EC_number="2.6.1.62"

/codon_start=1

/transl_table=11

/number=5

/evidence=experimental

/product="DAPA SYNTHASE"

/protein_id="CAA02327.1"

/db_xref="GI:2294847"

/translation="MTTDDLAFLPDRIHWPYTSMTPLPYVVSACECELLISGRF
LYDGSMMAAIHGYNHQLNAAKMSQIDAMSHWFGITTHAPATELCRKLIVAMPPOE
LECVLGSQSYAIVAKKMLQTYQAKGEARQRLTRNGYHEDTPEAMSVCPDRNL
MHSIMKGLVGENLFAPOSTRMDGEMDERDINVGARLMAHRHETAAVITIEIVQAGAC

BASE COUNT	1318 a	1552 c	1695 g	1307 t
stem_loop terminator				
RBS				
gene				
CDS				
	5098..5574			
	/standard_name="Orf1 RBS"			
	/gene="Orf1"			
	5098..5574			
	/gene="Orf1"			
	/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"			
	/codon_start=1			
	/transl_table=11			
	/number=6			
	/evidence=experimental			
	/product="PROTEIN"			
	/protein_id="CAO2328.1"			
	/db_xref="GI:2294848"			
	/translation="MKLISNDLRDGDCLKPRHFVNGMGYDGDNI SPHLAMDVPAGT			
	SYVMYCYRDPARPGSGMMHWVVVLLPADTVLRLVGGSGSLVAMPDGVLTQTRDFGKT			
	YGGAAPRKGETNRIETFLYHALDIERIDVDGSGAMVGFNVHFSLASITAMS"			
	5583..5605			
	5583..5644			
	/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL			
	TERMINATOR"			

BASE COUNT	1318 a	1552 c	1695 g	1307 t
ORIGIN				

ORIGIN

Query match	Score	DB	Length
100.0%	25	9	5872

```
Best Local Similarity 100.0%; Pred. NO. 0.00011;
Matches 25: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
```

MacLure's	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979	1978	1977	1976	1975	1974	1973	1972	1971	1970	1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	1959	1958	1957	1956	1955	1954	1953	1952	1951	1950	1949	1948	1947	1946	1945	1944	1943	1942	1941	1940	1939	1938	1937	1936	1935	1934	1933	1932	1931	1930	1929	1928	1927	1926	1925	1924	1923	1922	1921	1920	1919	1918	1917	1916	1915	1914	1913	1912	1911	1910	1909	1908	1907	1906	1905	1904	1903	1902	1901	1900	1899	1898	1897	1896	1895	1894	1893	1892	1891	1890	1889	1888	1887	1886	1885	1884	1883	1882	1881	1880	1879	1878	1877	1876	1875	1874	1873	1872	1871	1870	1869	1868	1867	1866	1865	1864	1863	1862	1861	1860	1859	1858	1857	1856	1855	1854	1853	1852	1851	1850	1849	1848	1847	1846	1845	1844	1843	1842	1841	1840	1839	1838	1837	1836	1835	1834	1833	1832	1831	1830	1829	1828	1827	1826	1825	1824	1823	1822	1821	1820	1819	1818	1817	1816	1815	1814	1813	1812	1811	1810	1809	1808	1807	1806	1805	1804	1803	1802	1801	1800	1799	1798	1797	1796	1795	1794	1793	1792	1791	1790	1789	1788	1787	1786	1785	1784	1783	1782	1781	1780	1779	1778	1777	1776	1775	1774	1773	1772	1771	1770	1769	1768	1767	1766	1765	1764	1763	1762	1761	1760	1759	1758	1757	1756	1755	1754	1753	1752	1751	1750	1749	1748	1747	1746	1745	1744	1743	1742	1741	1740	1739	1738	1737	1736	1735	1734	1733	1732	1731	1730	1729	1728	1727	1726	1725	1724	1723	1722	1721	1720	1719	1718	1717	1716	1715	1714	1713	1712	1711	1710	1709	1708	1707	1706	1705	1704	1703	1702	1701	1700	1699	1698	1697	1696	1695	1694	1693	1692	1691	1690	1689	1688	1687	1686	1685	1684	1683	1682	1681	1680	1679	1678	1677	1676	1675	1674	1673	1672	1671	1670	1669	1668	1667	1666	1665	1664	1663	1662	1661	1660	1659	1658	1657	1656	1655	1654	1653	1652	1651	1650	1649	1648	1647	1646	1645	1644	1643	1642	1641	1640	1639	1638	1637	1636	1635	1634	1633	1632	1631	1630	1629	1628	1627	1626	1625	1624	1623	1622	1621	1620	1619	1618	1617	1616	1615	1614	1613	1612	1611	1610	1609	1608	1607	1606	1605	1604	16
-----------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	----

1 ggatctgctgttgaagcgcagcag 25

178 CCATCTCTCTTGAAGCGCAGTAC 203

RESULT	7		PAT	05-MAR-1997
A38251				
LOCUS	A38251	5872 bp	DNA	
DEFINITION	Sequence 6 from Patent WO9408023.			
ACCESSION	A38251			
VERSION	A38251.1	GI:2294649		
KEYWORDS				
SOURCE				
ORANISM	Escherichia coli.			
	Escherichia coli.			

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;

COMMENT
other publication PL 308301 950727
other publication CA 214540 940614
other publication AU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9508809 950913
other publication FI 951547 950331
other publication JP 8501694T 960227

FEATURES	Location/Qualifiers
source	1. .5872

```

RBS      1141..1156      /standard_name="BIOF" RBS
gene      1154..2308      /gene="BIOF"
CDS       1154..2308

```

```

/gene="BioF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINALDARRADALRRRYPAVAGRWLVADPROYIN
SSNDYLGSHHPQITRAMOQAEQFGISGSGSGSYVHOALEELAEWLCYSA
LLFISGAANAVAIAMAKEDRIIARLSHLSILEASLSPQLRFRANDVTHLR
LLASPCGQOMVTEGVSHDGSAPLAEIOYQOHNGWLMVDDAGTGVIGQNG
SCHLQVKAPELLVTEGKFGVGAVALCSIVADYDLOFARHLIYSTMPAQOAL
RASLAVIRSEGDARREKLAALITFRAGVODLPFTLADSCSAIOPLIVGDSRALOL
AEKLRQGCWVTAIRPTVPAGTARLRLTLAAHEMODIDRLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYPVNGTDEVEKTVASCALLQAKAGRTGKPVASGS
EKPEKGRNSDALALONSSLOLDYATVNPYFAEPTSPHIIISQGRPIESLWAG
LRABQADWLVLEGAGWFTPLDFTFADWYTOQLPVILVGVKLCINHAMLTA
QVIOHAGLTLAGVANDVTPPKRAHAEYMTTLRMIPAPLGEIPWLAENPENAAATGK
YINAEVDASTLGFETSL"

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacatgctgtttgaagcagcag 25
|||||
Db 179 GGATCTGCTGTTTGAAGCGACGAG 203

```

RESULT 8
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
FEATURES
location/Qualifiers
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
promoter

```

```

-35_signal
-10_signal
RBS
gene
CDS

```

```

/function="PROMOTER PTAC"
/evidence=experimental
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/number=1
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPWTLSQVTELEKPLDLLFEAQOYHROHFDPROYVST
LISITGACPECKYCPQSSRYKTGLEERLMEVEVOYLESARAKAGSTPCGAM
KNPHERMDPYLEOMVOGKAMGLECMTLGSQOARLNAGIDYTNHLDPSPE
VGNITITTRYQERLDTLKVADAGIKVCSGIGVIGETVKORAGILLQANLPTPES
VFINALVKGTPLADNDVDAFDFIRIAARIMPTSYVRLSAGREONNEOQAMC
EMAGANSIFPGCKLLTPNPEKDLQLFRLLGINPQOTAVLAGDNEQORLEOALMT
PDDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/transl_table=11
/codon_start=1
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVYNKOALAAFGRAAAHYEQHADLQROSDALLMLPQRYT
HYLDAGCGPQWMSHMRERHQAQVTAALDSPMLVQAQKDAADHYLAGDIESPLATA
TFDELPMNLAVQWCGNLSTALRELYRVVRPGVAAFTTIVGSLPELHOAMQVDESR
HANRFLPNDLDEIOSLNGVYOHNIOTITLMPDDALSMRSLKIGAIHLEHGNPRL
TRSOIORQLALMPQOGRYPLTHLFLGIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE-8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDORHIMPTSTPLVYPVYVABGCELLISDGR
LVDSQSWMAIHYVNPOLNAAKSOIDMSHVMGITVHAPELACRLVADTPQ
LECVFLADSGSAVEYVNPOLNAAKSOIDMSHVMGITVHAPELACRLVADTPQ
MHSIMKGYLPENLFAPQOSRMDGEMDERMVGARFLMAHREHIAVITEPIVQAG
GMRMYHPEMLKRIKICDREGILLIDELITGFGRTGKLPACFHAELADPILGKAL
TGGMVLTATLTLTREVAFETISNGEAGCGMGPFPKNGNPLACAAANASIAESGMOO
OVDIEVOLTREOLAPADADMAVDVRYLGIYVETTHPVNMAALOKFVEQVWIRP
FKGLIYPPYIILPQOLRLTAVNNAVODETFEQ"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"

```

```

RBS
gene
CDS

```

```

/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRIVFNGMGYDDGDNISPLAMDVPAGTK
SFVVTCTDPAFTGSGMMHWVYNLPADTRVLPQCGSGLVAMPDGVLTQTRDEKTK
YDGAAPPKGETGHRVIFTVAHLDIERIDVDEGASGAMVGFMHFSLASASTAMFS"
stem_loop
terminator
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ggatctgctgttgaagcgagcag 25
        |||||||||||||||||||
Db      179  GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT  9
A93679   5872 bp      DNA      PAT      22-JAN-2000
LOCUS    Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679.1 GI:6741867
VERSION   A93679.1
KEYWORDS  Escherichia coli.
          Escherichia coli.
SOURCE    Escherichia coli.
ORGANISM  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS  Birch,O., and Brass,J.
TITLE    Biotechnological method of producing biotin
JOURNAL  Patent: EP 0798384-A 6 01-OCN-1997;
          LONZA AG (CH)
FEATURES  Location/Qualifiers
          source          1..5872
                        /organism="Escherichia coli"
                        /strain="DSM498"
                        /db_xref="taxon:562"
                        /clone="PBO30A15-9"
                        1141..1156
                        /standard_name="BIOF RBS"
                        1154..12308
                        /gene="BIOF"
                        1154..12308
                        /gene="BIOF"
                        /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                        /EC_number="2.3.1.47"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="KAPA SYNTHASE"
                        /protein_id="CAB69594.1"
                        /db_xref="GI:6741868"
                        /translation="MSWQEKINALLARRADALRRPYAOGCAGRLVADDDOYLNF
                        SSNDYGLSHHPOLIRAMQGAEOFGIGSGSGVSGT SYVHQLBEELAEKLGYSKA
                        LLEISGPAANOAVIAAMAKEDRIARLRLSHASILLEASLSPSOLRRFAINDVTLAR
                        LIASPCGOOMVVEGVFSMDGSAPIAEIQOYHNGWIMWDAGTGVIEGGRG
                        SCHWQKVPPELLVTFPGKFGVSGAVICSTVADYDLOFARHLIYSTSMPPAQOAL
                        RASLAVIRSDGDAARRKLAALITRFAGVODLPFTLADSCSAIOLPIYDNRBALQL
                        AEKLRQGCWVTALRPPTVAGTARLKLTLTAHEMDIDIRLLEVLHGNC"
RBS      3030..3045
        3030..3045
        /standard_name="BIOF RBS"
        3043..3753
        /gene="BIOF"
gene

```

```

CDS
3043..3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFTVGTDEVGKTVASCALLQAKAAGRTAGKPYASGS
EKPEEGLRNSDALIQRNSSLOLDYATVNPYTAEPSPHIIISAQEGRPTELSVMSAG
LRALEQADWVLVEGAGGWFPLSDFTFADWVQEOPLVILVGVVLCINAMTLA
OVIOHAGTLTAGWANDVTPPGKRAHAEYMTLITRMIAPLIGELIPVLAENPENAAATGK
YINLAFVASTIGFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ggatctgctgttgaagcgagcag 25
        |||||||||||||||||||
Db      179  GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT  10
AR101809 5872 bp      DNA      PAT      14-FEB-2001
LOCUS    Sequence 1 from patent US 6083712.
DEFINITION AR101809
ACCESSION AR101809
VERSION   AR101809.1 GI:12812607
KEYWORDS  Unknown.
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS  Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE    Biotechnological method of producing biotin
JOURNAL  Patent: US 6083712-A 1 04-JUL-2000;
          Location/Qualifiers
          source          1..5872
                        /organism="unknown"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ggatctgctgttgaagcgagcag 25
        |||||||||||||||||||
Db      179  GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT  11
AR101810 5872 bp      DNA      PAT      14-FEB-2001
LOCUS    Sequence 6 from patent US 6083712.
DEFINITION AR101810
ACCESSION AR101810
VERSION   AR101810.1 GI:12812608
KEYWORDS  Unknown.
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS  Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE    Biotechnological method of producing biotin
JOURNAL  Patent: US 6083712-A 6 04-JUL-2000;
          Location/Qualifiers

```

source 1..5872
 /organism="unknown"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaatgcgtttgaagcagcag 25
 |||
 Db 179 ggaatgcgtttgaagcagcag 203

RESULT 12
 AE000180
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 JOURNAL

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Submitted (13-0CT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30333 (e-mail: mark@ember.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and

its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
 source
 repeat_region
 gene
 CDS
 promoter
 gene
 CDS
 protein_bind
 gene
 CDS

Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
 complement(147..1430)
 /gene="yphc"
 /note="b0772"
 complement(147..1430)
 /gene="yphc"
 /function="orf; Not classified"
 /note="f427; 98 pct identical to fragment YBHC_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
 /codon_start=1
 /transl_table=11
 /product="putative peptidesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSRSRLALAFVTLTACSTPPDRSPDPTAGTSSRPTLSAFAONEDAOHYFASLTPGAAMNPSPITLPPQPPVNPACTOCYTHRTTAAVDAILIKRTKROYIAYMPEXCYTVYPAAPSGITLVTGKRPIDYKIGLSIDGMSFA DMRHDVNPGRKIMPEKPAWIMTDCQKRSQSTICVLSAVFSSONGLQQLNTIENT LGSVDAGNHPAVALRTDGDQVQINNVNITIGKQMTFFVYNSGVNRLTNPPTLVY NSYLEGDVIVSGCAVVDNTEFRVNVNSTRQDAVYFAFATLSNITYGFLAVNSRN AFGVDQVLQSLVDANTNGQVVIIRSAINEGFNTAKPMADAVISNRPFAGNTGSDV DNDLEIQRNLINDNTYNNRMVEYNNRGVSGSVAAEAK"
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582..2058)
 /gene="ydhb"
 /note="b0773"
 complement(1582..2058)
 /gene="ydhb"
 /function="orf; Unknown"
 /note="f158; 99 pct identical to YBHC_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRGDKLPHRHVPNGMGXGDNISPLIADDPYPAKTSFVVTGVDPAAPGSGGHWVYVLPADTPVLQGRSGGLVAMPDGYLQRTPTGTG YDGAAPRKGKHFRTFVHALDRIIDVDGASGAMVGFVHFHSLASITAFMS"
 complement(2108..2124)
 /note="central position to predicted promoter:85..5"
 /bound_moiety="Rnas predicted site"
 complement(2117..3406)
 /gene="b0774"
 /note="b0774"
 complement(2117..3406)
 /gene="b0774"
 /EC_number="2.6.1.62"
 /function="enzyme; Biosynthesis of cofactors, carriers: Biotin"
 /note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"

Query Match

/db_xref="GI:1786991"
 /translation="MTTDLAPDQRIHWPYTSMTSPDPVYVVAEGCELLISDGR
 LVDGMSWMAIIGHYNHNPOLIAAMKSOIDAMSHVMGGLTHAFALICRKLAMPQP
 LECVPLADSGSAVEAVEMKALIOYMOAKGARORFLFRNGYGDFTGMSACDDPNS
 MSLIMKGYLPENFAPAPOSRMGDEMDEMRMVGAFARMLAHEIAVILIEPVGAG
 GMMKMYHEMKRRTKICDRGILLINDELATGTGRTGKLFACCHADIDICLKGAL
 TGGTMTLSATLTTRVAETTSNEAGCFMHGPTFGNPLACAAANSIALILESGMOO
 OVADIEVLEQLOAPADAEVADVGLGIVETTHPVNMAALOKFVEGCVWLRP
 FGKLIYLPYIILPOOLRLTAAVNAVADDERFPCO"
 complement(2193..2221)
 /gene="bioA"
 /note="factor Sigma70; predicted +1 start at 807260"
 3411..3450
 /note="central position to bioB promoter: -20"
 /bound_moiety="BioB documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="BioB documented site"
 3413..3441
 /note="factor Sigma70; promoter bioB; documented +1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 3493..4533
 /gene="bioB"
 /note="b0775"
 3493..4533
 /gene="bioB"
 /EC_number="2.8.1.-"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346: 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /translation="1"
 /product="Biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPMTLSQVTELFEKPLDLLFEAQVHROHFDEROVAVST
 LLSIKTACPEDCKYCPOSSRYKTGLFAERLMEVOLESARAKAGSTRFCGAMV
 KNHFERMDPYLEOMVGVKAMGLACQMTLGLTSSOARLANAGLDYNNHLLPSPER
 YGNITTTTYOERDLTEKRYDAGIKYSGSLGIVGKXDRAGILLIOLANLPTPES
 VPINMLVAVKSTPLADNDVDADFIRITAVARIMPTSYRLSNGREOMEOQAMC
 FMAGANSIFYGCKLTTPNPEKDLQLFRKLGINPOQTAVLADNEDQRLQOALMT
 PDDEYNAAL"
 4530..5684
 /gene="bioF"
 /note="b0776"
 4530..5684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384: 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /translation="1"
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKINALDARRADALRRYPVAGAGRWLVADROLYN
 SNDYIGLSHNPOLIRANOGCAEFGICSGSGSYVVAQALFEELAEELGASRA
 LFIISGFRANQAVIAAMAKEDRIADRLSHSLILEASLSLSQIRRAHNDVTHLAR
 LIASPCGQGVVTEGVFSMDGSAFLAEIQVTOOHNGWMLVDDAHSTGYTGEGRG
 SCMLQVKKPELLVTFGKGVGGAALVCSSTVADYLLQFAHLLIYSTMPAQAOL

100.0%; Score 25; DB 1; Length 11022;

Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
 |||
 Db 3555 ggaatcgtctgttgaagcgcagcag 3579

RESULT 13

AE004192

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

/gene="VC1112"
 /note="similar to GB:J04423 SP:P12996 PID:145425
 PID:490219 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAFP94271.1"
 /db_xref="GI:9655583"
 /translation="MEVRHNTVAYEYKALLDPRMDLFEAQVIRLHHPINHOVST
 LLSIKGACGEDCKYCPQSAHRTIDYDKSRLEVERVLDAQAKANSSSTRCKGAA
 KNRKEDMLPKEMTIGRQDMGLTCTMLTLPQAOQLAQAGIDYNNHNDSP
 YGNIITRTYQDLDTLTSVRDAGKICSGGIIIGGESTNDAGLIVELANIPTREE
 VPINMILVKVGPLEQVDVDEPFEDFVRLIAVARIMKPSAVLSAGREKMDQALQ
 FMAGPASTIEGCKLTTPRPAEDSDMLFEKKGINREYQAOKPDEITENELDRVE
 VAARPAISDFEYPAAL"
 2533 . 3687
 /gene="VC1113"
 2533 . 3687
 /note="similar to GB:J04423 SP:P12998 PID:145426
 PID:490225 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAFP94272.1"
 /db_xref="GI:9655584"
 /translation="MTLRPSRIEOLANBREHRLTRRLDLSACQELIQORPTVY
 NESNNYLDALDAELVQAWQMTGLAPFGACASPMYGFSSRAHELEHLCMTLGE
 RAVLSSGSANQALLFTLEKEEDLLDDKLNHSLMAGMLSPATMKRKHNDESL
 ROLHROSGSYVALTEGVSFMSDDCARLQIATLVAKORAMLVDDAHGIGVIGEGACSS
 COAGIHPRELIVYTGKAFGLGAVALIQAIDVGHYLTQFARHHVSTAMPAPQAAGATL
 HAQMILQSWSRREODIELLACFEDQCRNIGFVATQPIKPMILGSDSLASHEL
 KQKGIWVASAIREPTEVGSARLRTITTAHQAIIRLSFOLKHYMERDT"
 3549 . 4487
 /gene="VC1114"
 3549 . 4487
 /note="similar to GB:J04423 SP:P12999 PID:145427
 PID:490221 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis protein B10c"
 /protein_id="AAFP94273.1"
 /db_xref="GI:9655585"
 /translation="MGECSTTHRACROCATFYHNGCAGYPSADTPLLKAVTKDGRG
 HSMSTATCELYELKDKSAIDQAFGAANYHOHNAFQDVLRLIOKMPSCGLRLV
 DLGCTGTFSALRLRGAQVYCADISHEMLDOARQDQSGMSYQLADAROLDPFA
 SGMVSSATLQWCEDELSLPLEIRIRVLKPHQOARLSTWLLDGSPLFELQAKRSVDHNRH
 INQFISINQVETIALAQCQSQHHLDIAITVYTAAPMLNRDLKIGIGAHVSGRSTGL
 ISRLRLAVEIAYOSFRMQGLVPATTQVCLGVTRH"
 4463 . 5179
 /gene="VC1115"
 4463 . 5179
 /note="similar to GB:S68059 SP:P13000 PID:490223
 PID:145428 PID:460901; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase"
 /protein_id="AAFP94274.1"
 /db_xref="GI:9655586"
 /translation="MFRGYSPMRNATFIACDTDVGKTVASKAIILOALATHNATIGY
 KPVYAGSKTEFGRNSDADLHLMKAADVDPYADVPAVLVPTSHIAKHENYTDI
 TALSLNLSKSRKQNAELVVEGAGMGWPTSDCLSTWQGRPLVILVAGIKGLCL
 SHAITAEVIRADGLELVGMIVANRNGTEHYAETIEHLEGRIGTGPKLAEITPYMRKAK
 ROELGKFTIIDLHLEPDTVA"
 complement(5261 . 5470)
 /gene="VC1116"
 complement(5261 . 5470)

Query Match	100.0%;	Score 25;	DB 1;	Length 12891;
Best Local Similarity	100.0%;	Pred. No. 0.00012;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
DB	1559	GGATCTGCTGTTTGAAGCGCAG	1583	
gene		/gene="VC1116"		
		/note="identified by Glimmer2; putative"		
		/codon_start=1		
		/transl_table=1		
		/product="hypothetical protein"		
		/protein_id="AAF94275.1"		
		/db_xref="GI:9655587"		
		/translation="MSRRNLTNKGCKHMAEAHSYQMAIDLCCHLGISPQEARQLG		
		EAEQTLQOOIVDTQISLGLSAEK"		
		5697. .6560		
		/gene="VC1117"		
		5697. .6560		
		/note="similar to GB:M58470; SP:P23894 PID:146412 GB:000094		
		PID:1736470; identified by sequence similarity; putative"		
		/codon_start=1		
		/transl_table=1		
		/product="heat shock protein HspX"		
		/protein_id="AAF94276.1"		
		/db_xref="GI:9655588"		
		/translation="MKRILFLATNIAVLVLSVLTNYAVTGNQPSLSGLLYMAF		
		VFGCGAFISLMSKSMALRSVGVDITPPRMEHMLLETVRQAGAGIGMPVA		
		YAPRMNATGAKRSDSLVAVSTGLNMTDEAEAVLAHEVSHIANGADVTTTLM		
		GVNVTVEPLSRFIANIYASRDEEGESNMNVYGVGVSMLEIVGLFASPTLTMYS		
		HEEFDVAGAAVLQVKKHIMALERLKMGQESHLEGSMMAGCITGCKRSLSELMTHNP		
		LEKRIALRNP"		
		6770. .7198		
		/gene="VC1118"		
		6770. .7198		
		/note="similar to PID:1653608; identified by sequence		
		similarity; putative"		
		/codon_start=1		
		/transl_table=1		
		/product="transcriptional regulator, putative"		
		/protein_id="AAF94277.1"		
		/db_xref="GI:9655589"		
		/translation="MSIMDYHVAQFYQQLDRSOLRLIETIYHDPVFEADAHRTEGH		
		DALYQPLNYQNMVTCFTIHEQYAVNEGAFVMTMLRPLAKGQYDVKGVSH		
		HFAEKGVTNRYEPLMGEMLYEQPLVGLQVIRAIKRLRGQ"		
		7195. .7920		
		/gene="VC1119"		
		7195. .7920		
		/note="similar to SP:P25970; identified by sequence		
		similarity; putative"		
		/codon_start=1		
		/transl_table=1		
		/product="oxidoreductase, short-chain		
		dehydrogenase/reductase family"		
		/protein_id="AAF94278.1"		
		/db_xref="GI:9655590"		
		/translation="MTAVFTIGATSGIGKQLADLYAKQGMQVIACGNSQVLSLHTQ		
		YANLEPLAFVDYTHDPRGTAKALADLCQPELMTINAGCEYIDGKMDVTLMARFEN		
		VLGAAYVEGILPHISGRHVAIVGSIASELAPRAVYGSRAAVYIARLTOLDPR		
		PLGEVTTIPGEVATPLDTRNPFAMNITVERAOGIKAGLARGVSQLYPFRFTW		
		LIRLGLPYAMQGRILYRLLKA"		
		7931. .9211		
		/gene="VC1120"		
		7931. .9211		
		/note="identified by Glimmer2; putative"		
		/codon_start=1		
		/transl_table=1		

RESULT 14
 LOCUS AE005258 13501 bp DNA BCT 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155.
 ACCESSION AE005258 AE005174
 VERSION AE005258.1 GI:12513751
 KEYWORDS
 SOURCE Escherichia coli O157:H7 EDL933.
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 AUTHORS 1 (bases 1 to 13501)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postell,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,J.,
 Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL Nature 409 (6819), 529-533 (2001)
 MEDLINE 21074935
 PUBMED 11206551

REFERENCE
 AUTHORS 2 (bases 1 to 13501)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postell,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,J.,
 Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 Location/Qualifiers
 1..13501
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 <1..7576
 /note="O-island #36: Region of the EDL933 chromosome not
 homologous to E. coli K-12 M61655; Cryptic prophage
 CP-933K; Includes one copy of the 13 bp direct repeat that
 flanks the prophage"
 66..665
 /gene="lomk"
 /note="20981"
 66..665
 /gene="lomk"
 /function="putative membrane; other or unknown (Phage or
 prophage Related)"
 /note="Residues 1 to 199 of 199 are 72.86 pct identical to
 residues 1 to 199 of 199 from Genpept 118 :
 g1175327891gb1AA63231.11AF151091_2 (AF151091) lom
 [prophage P-EB1a]"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein of prophage
 CP-933K"
 /protein_id="AAG55137.1"
 /db_xref="GI:12513752"
 /translation="MKRVCAIILSAICLAVSCVPAMASEHOSTLSAGILHASTDARG
 SDLNGINVKRYEFTDTLGLITFSYANAEDOKTHYSIDTRWHEDEVNRNWFSSVAG
 PSYRVNWFSAVMAGVAVSVSTFSGDYFRVDNKRKTHDYLITGSDDARYSNTSLAW
 GAGVQNPTESSVADVAVEXSGSGDWRTDGFIVGVCYKF"
 719..2041
 /gene="20982"
 719..2041
 /gene="20982"
 /function="putative structure; Structural component (Phage
 or Prophage Related)"

/note="Residues 164 to 440 of 440 are 68.79 pct identical
 to residues 381 to 645 of 645 from Genpept 118 :
 g1145854361gb1AAD25464.11AF125520.59 (AF125520) putative
 tail fiber protein [Bacteriophage 933W]"
 /transl_table=1
 /codon_start=1
 /product="putative tail component of prophage CP-933K"
 /protein_id="AAG55138.1"
 /db_xref="GI:12513753"
 /translation="MGMAAVOISGVLDGAKPKPQNTTOLKARNSTYVYVTVASE
 NPDEAGRSMDVEYGOYSVTLVPEFPPSHAGTTSYEDSQPGTLNDFLCAWEDDKR
 PEALRPEOMVEEARRHAEKKNAGCAETTSARMAGISASAKASAMADTSADSE
 SAROAAEASAAKKESEASSASAEAKASEISQATDALESKTAEASGAARDA
 TTSFEKARESAOSAEOSRIAEADVNRPVYVGPPEGPGEPGAPGPGKDGKE
 RGDTPGKATGERPGDGTGAGCGKCGKEGEGTGLGNMGPGKDGAPGAPGA
 PGPGKGTGAGVGTATGPGPGKDGEGTGTIRRLDPMRIETNSGMPGIDGALI
 TGLTFLEPKDITVOGMEFHLQVRRGGQWQDVKGLDEVSDTGRTE"
 2004..2312
 /gene="20984"
 2004..2312
 /gene="20984"
 /function="orf; Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 1 to 102 of 102 are 98.03 pct identical to
 residues 28 to 129 of 129 from Genpept 118 :
 g1145854371gb1AAD25465.11AF125520.60 (AF125520)
 hypothetical protein [Bacteriophage 933W]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55139.1"
 /db_xref="GI:12513754"
 /translation="MKNAVIOAELENDNMLIRKIMOSLCGGKRDHDCNGRSILQALR
 LGPADLSDENGLIPEDDRVTVOVILLADKKOICVVRPLQLLRADGTWENIGMK"
 2489..3469
 /gene="20985"
 2489..3469
 /gene="20985"
 /function="orf; Other or unknown (Phage or Prophage
 Related)"
 /note="Residues 13 to 326 of 326 are 55.73 pct identical
 to residues 18 to 331 of 336 from Genpept 118 :
 g1169603671gb1AA63327.11 (AF170116) hypothetical protein
 predicted by Glimmer [Salmonella typhimurium lr2]"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55140.1"
 /db_xref="GI:12513755"
 /translation="MLSPIRTFHNHNVNIYVSSPCQVSPAGKEVELKVIDEKPILF
 OMEFNPERRYKDEVPYNTKQHPYLDVNTAARRESRMIGIVDDGFSNOTKARS
 KLERDENVMYIYREDVDFSMYDRKLSIDYHDTICEORLTREDRETLMLLKELR
 EISKQDSLISWYAKRRNHAMDFERNLALKAGEIFPCYNTNRNHGISEGCTYLD
 ADMILTGKLGITVADGIGIMYVDRNRDENVNENSI1YVNRNHPALLEGLSEFMSKVD
 AHYVDYGKGVKKYFNTPLHNHNFCDPIEFNHPNIIIMNTSQTCSW"
 3530..4522
 /gene="20986"
 3530..4522
 /gene="20986"
 /function="orf; Other or unknown (Phage or Prophage
 Related)"
 /note="No significant matches"
 /codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55141.1"
 /db_xref="GI:12513756"
 /translation="MKIPSLOSNEFNSAPAGSAPFIAPRAENAYADVLDIGRIPL
 SAADLSNYEESVIRAVHDSRSLIDOHVDMIGNVLDALSRSGFRPAYSYGTHNEK
 VHIGCTIRNREYLEESSYKIDDIQSUTCNELIYKDYVGPRIPIPCAGENDPEPY
 VSFVAPDPTSEMCSSWQGLIETIHHVTSDDSGSDSNTLEGTETILARVQDELG
 MSVPDFKGYAEPEEREAHLRLNLRLROAMKHEENEFERIGTISDRYEASDFT

```

gene
5350. .6231
/gene="20989"
CDS
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/transl_table=11
/translation="MLSPSYNLCGCSNLSRLTSPDNVLSVRAAVSHNGACV
KGNRYRYVATDNKPCVTRSHSGCTNLHLRKGELISRIEVLNAPSVSAM
ERGLVHSNRDLPVDYAPPLPSVDYNRISVPCGNAYVADADAVKLM
FTTSQSEVTEVRCFNOYAGSARKIYGNNDIIGIMDKINGESLINSILPAO
AEHAIYDFRLEFGKILFYDTETNVLVPRANKFNPIDISSYNSDRSMSRSQIMQ
SYHGKODLISVLSKI"
6462. .7160
/gene="20990"
CDS
6462. .7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/transl_table=11
/translation="MRPTSLNVLHOSRSSMSDDIESLKASVOMIKNPOLRF
QGDHNIYQOITFALDKIGSTETGRLVNLNIESISRLKSEVTHLNSRLGYMARD
IDENHGCTGDFHCNLAWEYPCGEGISVVDHATITVEHLLHVNHLNGELKYES
SRMSQKYSPLLEEARVTLGAFSEVLSSENKFEHIGMPRTSTPPXDSALITHDNT
VSLGQGVRLHPLI"
complement(7668. .8144)
/gene="20992"
/note="20992"
complement(7668. .8144)
/gene="20992"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to residues 1 to 158 of 158 from Escherichia coli K-12 strain MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/transl_table=11
/translation="MKLISNDLRDGLRPHRVHNGMGYDGNISPHLAMDVPAGTK
SFVTCYDPAIPGSGMMHVVNLPADRVLPQGSGLVAMPDGLTRDPGKTG
YDGAAPKGETHRIYFTVALDVERLDVDEGASGAMVGFVHFHSLASISRAMPS"
complement(8203. .9492)
/gene="20993"
/note="20993"
complement(8203. .9492)
CDS
complement(8203. .9492)
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggaatcgtctgttgaagcgaagcag 25
|||||
Db 9641 GGATCTGCTGTTGAAGCGACGACGAG 9665
|||||

RESULT 15
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BAO00007
ACCESSION AP002553.1 GI:13360211
VERSION

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

```

Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
 Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

1 (sites)
 Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and Shinagawa, H.
 Complete nucleotide sequence of the prophage V12-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
 Genes Genet. Syst. 74 (5), 227-239 (1999)
 20198780

2 (sites)
 Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H., Hayashi, T.
 Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
 20557356

3 (sites)
 Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and Shinagawa, H.
 Complete nucleotide sequence of the prophage V11-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)
 20564182

4 (sites)
 Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shibata, T., Hattori, M., and Shinagawa, H.
 Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)
 21156231

5 (bases 1 to 297816)
 Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H., and Hayashi, T.
 Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
 genome project.
 Location/Qualifiers
 1. 297816
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79. .1245
 /gene="ECs0753"
 79. .1245
 /gene="ECs0753"
 /note="similar to SUCC_ECOLI g11786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="succinyl-CoA synthetase beta subunit"
 /protein_id="BA834176.1"
 /db_xref="GI:13360212"

gene
CDS
/translation="MNIHEYOAKOLFAPVGLPAPYVACTTPREAEFASIKGAPW
VACQVHAGRGKAGCVKVVNSKEDIIRAFENNLKRLVYTQTDANGGVYNOLIVBAT
DIABELLYGAVNDRSSRRVAVSTEGVEIEKVAETPHLHKVALDPLVPMYOG
RIAFRLGLEKLVQOFTKIFMGLATIFLELVALEIENPLVITKQSDLICDGLGA
DGNALFROPDLREMDQSEDEPREAOQAMELYNALGNICMYAGALMGTDIA
KLHGEPAFLDVGCGATKERYTEAFKILISDDPKAVLVNPGIVRCDLIADIIIG
AAVEGVNPPVVRLEENNAELGAKKLADSGNLITAAAGLIDAAQVVAANEK"
1245..2114
/gene="ECS0754"
1245..2114
/gene="ECS0754"
/note="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILDKNTKYICGFTSGCTSPHSEQAINGTKMVGCTPCKG
GTHLGLPENTVREAVATGATVAVIYVAPFCDSILIEAIDAGIKILITTEGIP
LMLTVKVLDAVGVMIGPNCVITPBECKIGIOPGHIKPGVGVISRSGLTVE
AVKQTDYCGSCVIGIDPIPGSNFTDILEMEKDPOTEAIYVIGETIGSABEEA
AAVKEHVKKPVVGYIAGVTAPKCKRMGAGAIAGGKTADKFALEAACVKTIRS
LADIGFALKTVLK"
2781..3686
/gene="ECS0755"
2781..3686
/gene="ECS0755"
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g11417043|sp|P32064|GCVL_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELVTFEVVARHESYTRAEEALITQSAVROVSAL
EEFLRLPLFSHKKRIFLNDGKYLGIVKELNKLERTNTIMTQPVQVTELVN
PFSITMLIPNHEFTKLHPDILVNIHSLANNGDIFNREDAVIMRENCAVAEEY
LPEETLIPVCSGSLAMSDOKLSVAELLLELPLDHOSTRITGMEEMFALSGVSPVYN
NGRPDLISMLIAVANSNGVALPRFALQHDLDSDMWIIPCDVPIRTGNRIIMTQW
EKSDSPHLQOFRFELAKSVQEM"
complement(3720..4322)
/gene="ECS0756"
complement(3720..4322)
/gene="ECS0756"
/note="probable cob(1)lamain adenosyltransferase, similar
to cob(1)lamain adenosyltransferases (cortinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P13040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)lamain adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEERHVROROKIKRQVDTVAVAITEKKGILIVFTGNGK
KSTRAAGTVTRAVGHRKTVGVAOYIKGMDNGRYNLLQPLGVEFHIMGCTFMTQNR
QADIDAKEWSESKRMLADKRYDLVLLDTYMLVHYLDLEEVYIASLONRPAQOSV
IVRGCHSOLIKMADTSEIRPVKHAFFNGIOAQDGM"
complement(4332..5984)
/gene="ECS0757"
complement(4332..5984)
/gene="ECS0757"
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g111205981|sp|P00923|FUMA_ECOLI percent identity 68 in 545

aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFIMQEPFLNKGDTETLISPDHITVTELDGEVETIAPE
ALTLARATFEASFPLRSARLQOVAIILNDQASNDKYALDILPMNESAQVLP
NCQDTFATTVASKGQOITWGNDAEALSKGISTFOENNLRSQNPDMTREVNTQ
TNLPAQIDISAVNAGDEYHFLCVNGGGSANKAAVQETKSLDPEKTLAFLEKMSL
GTACPPHYHIAFVYGLSADQTLIAKIASKYYDNPSTGNEGQAFDIEKXVL
EASQDFGIAQFGKYPFAHDRIYRLPRHGSCPIAMALSCADRNIAKINLGLWL
EKLEHNGOYTPASLRENNAOHYOLDNLRPMVDLRLPGTFSVLSGPVAVAR
DIARAKTKARLDSEPMPEYLNHVIYVYAGPATRPMACGSLGPTGSGMDYIDTF
OAAQGSVLMSKSGRSQVVDACHKHGFFNGSIGGAALLAQDYKASLSLCLPELG
MEAVMMMEVENLPAFILLVDKGNFESQFEQHRKASCAPAGH"
complement(6092..7372)
/gene="ECS0758"
complement(6092..7372)
/gene="ECS0758"
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671|sp|P24943|GLTP_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKRSITPMILLALVYGMIIIGVYINNPASPEFTAKVIAOEISFT
TFLUKLIMITIAPLVSTLVYVIGIAKKMGDARLGGIFSKTULFPLICASLISALGILTV
NPFMPGTINFVAHGAETTVAAPEFTLVKFIHAPPTSIVDMANNEIDLIQYVESI
FLGCSLTAIEKGSALVHALDLSLHAMIKLGYVLPAPLVEFALISALLAERGLAV
VSAGIFMEGEFFMLLMVLIGLAIYVGCIRLRALSEPALLAFITSSSEAPR
GLEKLEQFGVSPKIASFVLPIGYSFNISGMAYCAPATYVIAOACNIIHISIGQITM
LIIIMLTSKMGAVPRASVVYIAATLQNFNPEAGLILMLGVDFELMGSRATVWMS
AMGAAMVRMEGHEHDEGCRKALKPNSNALP"
complement(7533..7850)
/gene="ECS0759"
complement(7533..7850)
/gene="ECS0759"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.0% Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggaatcgcgtgttgaaagcagcag 25
DB 92373 GGAATCTGCTGTTTGAAGCGCAGCAG 92397
|||||
RESULT 16
LOCUS ECO10B 128 bp DNA BCT 03-JAN-1995
DEFINITION Escherichia coli biotin (biob) gene, early terminator region.
ACCESSION M27731.1
VERSION M27731.1 GI:341755
KEYWORDS biob gene; biotin.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 128)
AUTHORS Nath,S.K.

Query Match	Best Local Similarity	84.0%; Score 21; DB 2; Length 8227;
Matches	21; Conservative	0; Mismatches
QY 5	ctgctgtttgaagcgcagcag 25	
Db 4095	CTGCTGTTTGAAGCGCAGCAG 4115	
LOCUS	EBH10TOP2	1041 bp DNA BCT 04-NOV-1996

DEFINITION	VERSION	KEYWORDS	SEGMENT	SOURCE	ORGANISM
Erwinia herbicola biotin synthetase (biob) gene, complete cds.	U38648	2 of 2			
U38648	038648.1	GI:1228112			
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
Source					
gene					
CDS					
BASE COUNT					
ORIGIN					
Query Match					
Best Local Similarity					
Matches					
QY					
Db					
RESULT 19					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					

PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIRONO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12N1/00, C12N1/20, C12P13/18, C12N1/20, C12N1/19, C12P13/18, PC
 C12R1/19)
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH Key Location/Qualifiers
 FT
 CDS 42..1079
 /product="biotin synthetase".
 Location/Qualifiers
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN

Query Match 72.0% Score 18; DB 10; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ggatctgctgttgaagc 18
 |||||||
 Db 104 ggatctgctgttgaagc 121

RESULT 20
 AF281080/c 1614 bp mRNA VRT 01-OCT-2000
 LOCUS Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
 DEFINITION complete cds.
 AF281080
 ACCESSION AF281080.1 GI:10442649
 VERSION
 KEYWORDS
 SOURCE African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1614)
 Small, E.M., Vokes, S.A., Garlick, R.J., Li, D. and Krieg, P.A.
 Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
 Mech. Dev. 96 (2), 259-262 (2000)
 2 (bases 1 to 1614)
 Small, E.M., Vokes, S.A. and Krieg, P.A.
 Direct Submission
 Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
 Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
 85724, USA

FEATURES
 source
 1..1614
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /note="expressed in ventral forebrain, lungbud primordia
 and thyroid anlage"
 1..1614
 /gene="Nkx2-1"
 162..1205
 /gene="Nkx2-1"
 /note="thyroid transcription factor; TTF-1; T/ebp; chicken
 and mouse Nkx2-1 ortholog; contains homeodomain"
 /codon_start=1
 /product="homeobox transcription factor Nkx2-1"
 /protein_id="AA017405.1"
 /db_xref="GI:10442650"
 /translation="MSMSPKHTTPSVSDILSPLESYKKVAMEGAGIGAPLTAAAYR

OSVSAQSMQOHMGHNGCPVSAHYMTAGVPOLSHTTMGYCNGLNGLMSLEPP
 YQUTKANSATGAYGAYNPDRSTISRFMGPGGNNMGLGMSLGVGSMPLQA
 TPRKRRVLESOAYVELRRFRKQKQYASAPREHLSAMHILTPVOVLTWONHRYKM
 KRQAKDKASQOOIOQDNSSCOQOQSRVAVLVKDGKPCQAGSNFTPAALQSHQOO
 TATATITTINGLGPQHSHQOTNSAGSPDLVPHSNSSPSSLNQVTSLSHLNSSSDYGS
 AMSCSTLLVGRW"
 BASE COUNT 475 a 407 c 399 g 333 t
 ORIGIN

Query Match 68.0% Score 17; DB 8; Length 1614;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 tcgcgttgaagc 20
 |||||||
 Db 717 tcgcgttgaagcgc 701

RESULT 21
 AE001220 14268 bp DNA BCT 16-JUL-1998
 LOCUS Treponema pallidum section 36 of 87 of the complete genome.
 DEFINITION AE001220 AE000520
 ACCESSION AE001220.1 GI:3322705
 VERSION
 KEYWORDS
 SOURCE Treponema pallidum.
 ORGANISM Treponema pallidum.
 Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

REFERENCE 1 (bases 1 to 14268)
 Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,
 Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,
 Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,
 Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
 Uitterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,
 Weidman, J., Smith, H.O. and Venter, J.C.
 Venter, J.C. et al.
 Complete genome sequence of Treponema pallidum, the syphilis
 spirochete
 Science 281 (5375), 375-388 (1998)
 JOURNAL 98332770
 MEDLINE 2 (bases 1 to 14268)
 REFERENCE
 AUTHORS Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,
 Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,
 Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
 Uitterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,
 Weidman, J., Smith, H.O. and Venter, J.C.
 Direct Submission
 Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
 source
 1..14268
 /organism="Treponema pallidum"
 /db_xref="taxon:160"
 98..2149
 /gene="TP0421"
 98..2149
 /gene="TP0421"
 /note="similar to GB:AE000783 percent identity: 27.26;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AAC65409.1"
 /db_xref="GI:3322706"
 /translation="MRGWCISRSRYCAGALLLPAVGAQEHAPPVRYDALAGEF
 RGVLYAYRGAFNDAIOLFERRALSAPRNLIILEWIGNAYRSGISGAALHONGARD
 LGYGMALRNKITEVVOORDPAPDSADALHIFESSESPHAYRRTGLTFRRLTSCALAD
 GTWMSAYVSGNELLRFDVNGRVTARTGPEGRPDVYOTSGDLVSEPARSDSIC
 RLTKRGREPLKSPGKGGVGVOLIGPOFLADKIDNIYVDFGNARAVPADGAPLH
 FGKSAHFEGFSAGGIATIDELVYVADAIGAIIVFDGANVRLVAEGTLKHVES
 VRAWNGRLVSLPEVNVVDVGLASLHTIARIGNAPIRLTAAPVDPANGVLLADKYNE

RLEIVSRISLGLGLFVHFERVHADRPSPVNVSVHRECPVVGLOVSNFLTEEH
 RPVSEMRLOGAAYLNDTCSSLSLYDRSPATQERRLTYRVIGELAQVREGCTLSYS
 AAQSPVLEGTFTLAOLQAOPLRLAKSSSQWRFDLGLAGTLVADYKRAVFLSL
 GELQTPDAOYGVLDLAAYLTNNETSPVLOLQRTVLPVELGYLVQKTRGAVSVYAO
 EBLGALVAVHVSAGAGTYALSRLPTDFGRAVMPVEAEVRLSRSGRDETGYPAL
 E
 2133. 3158
 /gene="TP0422"
 2133. 3158
 /gene="TP0422"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0422"
 /protein_id="AAC65418.1"
 /db_xref="GI:3322715"
 /translation="MRRLSDRLRGMTGKGSMAVRLCYLSTQVGLAEPALRLAGW
 ENATPVAIAPTADAMRFSIMLVHKPYAVYERVAVSLSLSDARADVLPPPAOK
 AERVLVLSFEGRVVQPLVLCRMKDALFCSPFQVSYGTRSGSYSSFOITPRTDPL
 FEFWLEEGSAELRAVSSOAGELNAFFETANFRFRYWRDDALAPAOIRAYFTADD
 GVTYEPLOYVYOGAAVYSCITGRSRVVRNFOGTVEVRSIDGSKRLMLRRKGAQPGS
 YAVGAVYPROGFLDESALALRVGQFLAIGEPFLRKSAVHDVTAFIDAHMNRKAPRA
 PLVPDF"
 3183. 3989
 /gene="TP0423"
 3183. 3989
 /gene="TP0423"
 /note="similar to GB:AE000783 percent identity: 39.66;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65410.1"
 /db_xref="GI:3322707"
 /translation="MRVFSWMKOKDKGERERAGALOGACIGTPPEHROVFAVYGRSGS
 GKSFAOIVMRFRDIAFVYIDGLIREDKITNGRSKORITLAIIRVALPMTNHR
 VVARVITYFLCSQNKVLLITSEMKVRIALRVGLPERGHITRIEDIASSEELACA
 LTRHVEGHVIVPVSQEVKRSYPIETFERILULRRGAGERTGRMAHALHHEGLKR
 ACSGAHVAHEKSIVRPPSCNLRVVAATEGQDASPVVPPHEELAQO"
 4192. 4890
 /gene="TP0424"
 4192. 4890
 /gene="TP0424"
 /note="similar to GP:1510115 percent identity: 37.59;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit E, putative"
 /protein_id="AAC65411.1"
 /db_xref="GI:3322708"
 /translation="MLGESREAEIRIVRAAREAEIRIVRAAREAEIRISSIALSQ
 ASRNVLSFQDSVTRSLRAITSMETQAQADAGLRLIPRVYSAMVQAEGDLELILS
 PADLRLLEGVFCALGEOISAGVELRSDCLAGFRIVPAEGSYDFSAVAVOLFS
 SYVASARVAEVLILRRSCDVFLILPYNATISSMRGASFALSGFFGLCATPLGASGC
 CGMKLYVSRAGA"
 4868. 5128
 /gene="TP0425"
 4868. 5128
 /gene="TP0425"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0425"
 /protein_id="AAC65419.1"
 /db_xref="GI:3322716"
 /translation="MCPVPARDSTGSRVLDPMGEFFYALRCALARARARLRNDPVE
 VROCVAQESVAVARAVASLDDPLERAGVFEERALLTCRGV"
 5270. 7039
 /gene="TP0426"
 5270. 7039
 /gene="TP0426"
 /gene="TP0426"

/note="similar to GB:AE000783 percent identity: 63.40;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit A (atpA-1)"
 /protein_id="AAC65412.1"
 /db_xref="GI:3322709"
 /translation="MTQTKGISAVNCGMVSYTFEGVSLNEGYVHVCNARLKEII
 RVGRERAOLOVFEITRGVSGDVEPEFDLLSVELGGLGOYVDGLONPLPLAEKV
 GFLERGVNLPALSRITSEMFPTPRVSYGEYRVKDVIGTIPESALKHRIIVPRHMS
 YEVEFIQTACTYRVHDIARVDAQGHEHETLWAFRVPKRVPHCTAERIKPTPELVY
 SIPTIDYFPVAKGTICIPGPGAGTVDLQHSIRNADVDVYVIAACGAEVEET
 LREPDLDTPRTGSLMERTVIVCNTSMVAHSEASVGTGLAEYRQMGDLVLL
 ADSTRMAQALREMSGRLEIPGEAPALYESGTLIAFYERAGVRLRSGESVYTG
 CTSPAGCNPEEPYQATLKVACAFHGLSPRESDARYPAVHPLDSKSYPSYDARA
 VAYGRSLFRGAEEVOMRVYGEGRSMEPLVYLKGSFLDSYTLDOONSPTYDSAPV
 VAROKHCYAVIMRVLSVLAPESKDDARAFVSKUGHFIDYNCAMNSEAFEREKEI
 RAFLOGSTKIDSEABGIIIRGME"
 7043. 8335
 /gene="TP0427"
 7043. 8335
 /gene="TP0427"
 /note="similar to GB:AE000783 percent identity: 72.96;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit B (atpB-1)"
 /protein_id="AAC65413.1"
 /db_xref="GI:3322710"
 /translation="MYRYEKIESINGSVITVRAGVYHGEARVKTAFQDSIAEVNK
 LEGDVLSQVPAKRGISTGSEVFLRGMLVNSDYDLRIFNGSGVPNDGPALKE
 GRVLEGPSVNAKRVIAIRMIIRGITMIDVFLVYSOKRPISSSGEPYNELARI
 AMQAEQVYIVGGGCIKHDXLYFKSALFEAGLSRAVLVHTASDPTVGLVAPDMC
 LVAABQRLKGRDVLVLTDMTNPADMKELAIIOEVPYSKRGDGLYSQIASRYK
 LAVDDQAGSVTILAVTTMPGDVTHVPDNTGITYEOLFUKNRIEIPFSSLSRKN
 VNGETRADHRLALNMNMLKYAAYDALEKRSMGFAMSEWDEKLLTYGGLFEADLMDS
 VNPLEEAFDGMKILRCFPEETGIRSDLINYMPA"
 8347. 8967
 /gene="TP0428"
 8347. 8967
 /gene="TP0428"
 /note="similar to GB:AE000783 percent identity: 45.69;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit D (atpD-1)"
 /protein_id="AAC65414.1"
 /db_xref="GI:3322711"
 /translation="MAVRLTKNELRKQKTELTPFRRLPTQLKKOOLVYAIRVEAR
 ASEVSRADALRODOVQWVAVESGVDARLLVQVEKKTYASIAGRICFLEVHR
 ACPPDYLRVPLMDPTAVRKMOEVLISDLEHAYVLDQOALLGALRTTQGVNLFKV
 KIPETRACIKRITYILGDOVAAVYVRGMSKNLVDSRMQEDTR"
 8964. 10832
 /gene="TP0429"
 8964. 10832
 /gene="TP0429"
 /note="similar to GB:L77117 SP-Q57675 PID:1590960 percent
 identity: 25.19; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit I (atpI-1)"
 /protein_id="AAC65415.1"
 /db_xref="GI:3322712"
 /translation="MIVMKVYTLVLVGSDESRISQALRSFGANVHOLRECAQSLAE
 LHALDARCVQALVTAQOTKNVTRGECRAGOVENAEEVQIVRHSRVLEAOK
 IAOCIJALHERCEPMDPDAVRAALAQGHILIPVEISERSYRCLPELOTLCLARIG
 GLVRCVLAQDKRGPLSPADARALDLVDVSPADFLVRLQDLRECAITLQRLYAFS
 YOGARALRAOKCAADIEFERVHLISWVSVDVQWREGTTLIAHSGYLPVSRVAFS
 ECARKEANAAYCVDMPEDPVPQOLRNRRWNLSPLMNFGLTGVGVEVVISGPFIL
 FGVGFSTIFADAGVAVTLVSLGCIYLSKRAHVASPAMCIGIYLCITLMVNCALY
 CMNFGVPOYVAVSLARIAVETISGFADAQRNNKMHVCFGLGLHLCGLHLYVR
 TFRSLVLAIEFGSLIMLGMYVYVNLIVDERYPLTGMIVGSIAGVLFNFITVNR

VSYSVADSMKNNVINALGIVNPAVMSYIRLAVAGLACGASIVATNENTHPLFAN
FLATGLIVLFGHGLNVTMSIIVHGVNLTLFSNHGLMTGIRYTPFRER"

Query Match 68.0%; Score 17; DB 1; Length 14268;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gctgttgaagcagc 23
|||||
DB 8188 GCTGTTGAAGCGCAGC 8204

RESULT 22
LOCUS AX064067 924 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 349 from Patent WO0100843.
ACCESSION AX064067
VERSION AX064067.1 GI:12541779
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriales; Corynebacteriaceae;
Corynebacterium
1 (bases 1 to 924)
Pompejus, M., Kroege, B., Schroeder, H., Zeider, O. and Haberhauser, G.
corynebacterium glutamicum genes encoding metabolic pathway
proteins
Patent: WO 0100843-A 349 04-JAN-2001;

FEATURES
Source
Location/Qualifiers
1..924
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..904
/note="RXA01757"
/codon_start=1
/transl_table=1
/protein_id="CAC25274.1"
/db_xref="GI:12541780"
/translation="MPTASPIVYVVGAGTSGITATOLLDRAGLIKCPKSPKRG
RAVSVOOSDIFLGAATWPNLPLVQOLVNNIGLSTFQALTEGDLAFELVDA
RGNPDAASGRFQAGASSALGLAQIKRQVLELGPVHLSDEDEIVYKSKQIVR
AKHVIIVPAPALAEILIGFLIDLPADVRAAHPOHILAVNMWAKEKYLTPQAA
SAGF GHELFQOPLGHRIHMASTEVAFCGHEGAVRAGIQALQGFNLKS"

BASE COUNT 214 a 252 c 239 g 219 t
ORIGIN

Query Match 64.0%; Score 16; DB 9; Length 924;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtgttgaagcagc 23
|||||
DB 884 CTTGTTGAAGCGCAGC 869

RESULT 23
LOCUS YL1012084 6887 bp DNA PLN 23-SEP-1999
DEFINITION Yarrowia lipolytica pex10 gene for peroxin 10.
ACCESSION AJ012084
VERSION AJ012084.1 GI:5921565
KEYWORDS peroxin 10; pex10 gene.
SOURCE Yarrowia lipolytica
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodascaceae; Yarrowia.
1 (bases 1 to 6887)
REFERENCE Le Bail, M.T., Gallard, C. and Nicaud, J.M.J.
Pex10 is essential for peroxisome biogenesis in Yarrowia lipolytica
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 6887)
AUTHORS Nicaud, J.M.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
Microorganismes, INRA-CNRS, BP 01, Tiverval-Grignon, F78850, FRANCE
FEATURES
Source
Location/Qualifiers
1..6887
/organism="Yarrowia lipolytica"
/strain="W29"
/db_xref="taxon:4952"
complement(2293..3357)
/gene="pex10"
complement(2293..3357)
/gene="pex10"
/function="essential for peroxisome biogenesis"
/codon_start=1
/evidence="experimental"
/product="peroxin 10"
/protein_id="CAB56484.1"
/db_xref="GI:5921566"
/translation="MSDNTTIKKPIRPKPIRTERLPYGAAEIIRANOKHFPESVLE
QHLYVFLQKMGVREIHQYKELETPASFAVYGLCTIGVSKTIGREYTNLMYTRDR
ALPGVRRRGVLSNLTLPYLPYLRVYMGKLRANKIMREYHLYVEDPVPSPETWKR
VITFVKNKDFKFTALEGFTALIAIPVYKSYOLSKRINMRYVGRDKNPRIG
YEMGLILFARATSFYOTGRYTGALLLEKSYEKAGKEDEKAIVPKKSSITPLE
DEGETEDKIDLEDRLKFLPEASRACLTCLISTISAPACTPCGHFCWDCISBVR
KECPYKQGVNEONLPIR"

gene
CDS

BASE COUNT 1793 a 1841 c 1485 g 1768 t
ORIGIN

Query Match 64.0%; Score 16; DB 15; Length 6887;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gctgttgaagc 17
|||||
DB 4051 GATCTGCTGTTGAGC 4066

RESULT 24
LOCUS AE002054 11823 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans RI section 191 of 229 of the complete
chromosome 1
ACCESSION AE002054 AE000513
VERSION AE002054.1 GI:6460010
KEYWORDS
SOURCE
ORGANISM
Deinococcus radiodurans.
Deinococcus radiodurans
Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
1 (bases 1 to 11823)
REFERENCE White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
Science 286 (5444), 1571-1577 (1999)
2 (bases 1 to 11823)
REFERENCE White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers

TITLE
JOURNAL
FEATURES

CDS

7253..8008
/gene="DR2216"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1766.1"
/db_xref="GI:6460021"
/translation="MOLVQDRHGHTLRTHALVHTFSQSGDGLRETRMDGRFTV
EALISGGVLELPAGFYSVGOVASEQOQGGFLRAIATAGTESAFATLISGH
RWTEPRRHFIIEATGESGDMILDDRAFYLAODMOLGTHHSGSALSGGM
QPKLSGAGLFEVSEVPVSEVOYIELDRTVVDMDMLMYSASLRVELRLVRLR
NAARSGGIVFVLSGKGVLPFTHALSGASV"
9265..10326
/gene="DR2218"
9265..10326
/gene="DR2218"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1767.1"
/db_xref="GI:6460022"

Query Match 64.0%; Score 16; DB 1; Length 11823;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgtttgaagcgacg 23
|||||
Db 3308 CTGTTGAGCGCAGC 3323

RESULT 25

LOCUS AE004026 13856 bp DNA BCT 06-JUL-2000
DEFINITION Xylella fastidiosa, section 172 of 229 of the complete genome.
ACCESSION AE004026 AE003849
VERSION AE004026.1 GI:9107228
KEYWORDS
SOURCE
ORGANISM Xylella fastidiosa.
Xylella fastidiosa
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

REFERENCE 1 (bases 1 to 13856)
AUTHORS Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P.,
Terenzi, M.F., Truffi, D., Tsai, S.M., Tsuhako, M.H., Vallada, H., Van
Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M.,
Weidens, J., and Setubal, J.C.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
Nature 406 (6792), 151-157 (2000)
20365771

JOURNAL MEDLINE REFERENCE AUTHORS

2 (bases 1 to 13856)
Simpson, A.J.G., Reinach, F.C., Artuda, P., Abreu, F.A., Acencio, M.,
Alvarenga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S.,
Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Britones, M.R.S.,
Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carrato, D.M.,
Carrer, H., Colauco, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J.S.,
Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Ferreira, A.J.S.,
Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
Gomes, S.L., Gruber, A., Ho, P.L., Hohnsbein, J.D., Jungueira, M.L.,
Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,
Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,
Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L.,
Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,
Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G.,
Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C.,

REMARK FEATURES source

Palmeri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira
Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V.,
de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Silva
Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de
Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M.,
Tsuhako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S.,
Vettore, A.L., Zago, M.A., Zatz, M., Weidens, J., and Setubal, J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

Location/Qualifiers
1..13856
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(195..803)
/gene="XF2101"
complement(195..803)
/gene="XF2101"
/note="similar to GII1652989 (percent identity: 30 %/query
alignment coverage: 87.6 %/subject alignment coverage:
67.6 %) identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAPI1766.1"
/db_xref="GI:9107229"
/translation="MSRCITGIRMMQNICVAVVLPFWLVAAGVIWIGQDQOASS
DVIIVLGAADAKPSVFERIRHGLYVQGVAPPLITGGGSSSAFASQVAVR
RYALRDVPPRALILETVSSRTROLNLOARLMVYGLRAIVVSDPLHARALRLCR
ELIDIALASATPSTRFSLRTMRFFLLQEVYFFHRDLVQAN"
complement(800..1042)
/gene="XF2102"
complement(800..1042)
/gene="XF2102"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1767.1"
/db_xref="GI:9107230"
/translation="MARVIMYSSFALSVCIGVLTGMLLHAYVFGGIGGEPFAL
CCDVAWLCHVDISARLSGCCSAGFFICGVSLVVS"
1397..1732
/gene="XF2103"
1397..1732
/gene="XF2103"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1768.1"
/db_xref="GI:9107231"
/translation="MKARLPLILINSATGATVYNGYHGPVTEHYHPNYY
PHGGMHDLNPYIPYPRDFYRTHILHPYQWRPHHDSRPLRPKQPIKQEPNIE
MGLRDKQPI"
1914..2024
/gene="XF2104"
1914..2024
/gene="XF2104"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1769.1"
/db_xref="GI:9107232"
/translation="NNIADQDMLICSEILMKICPHCVDTLKQRLAR"


```

gene      complement(2067. .2225)
CDS       /gene="XF2105"
          /complement(2067. .2225)
          /note="hypothetical protein; identified by sequence
          similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84904.1"
          /db_xref="GI:9107233"
          /translation="MAAWSREAMMEFYFCVIERNALFRVGYFVAVDVVAPAAALRL
          OARSGLE"
          2271. .4160
          /gene="XF2106"
          2271. .4160
          /note="similar to SPIP25756 (percent identity: 65 %/query
          alignment coverage: 98.6 %/subject alignment coverage:
          98.4 %); identified by sequence similarity; putative; ORF
          located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="glucose inhibited division protein A"
          /protein_id="AAF84905.1"
          /db_xref="GI:9107234"
          /translation="MTDSFYGYDVIVIGAGHAGSHALAAATGANTLLTNHNETTG
          AMSCNPAIGIGKGLVKEIDALGAMMAHADAQIOWRTNASKGPAVRVRCOAR
          ALYRAIRIOMIGGARLNIPOAEVDLLFEGPTVCGALITHTGKAPAVVLTGTL
          AKGTHIGPTOYAGRAGDPPATMLAARLERFVYARLKTGPPRIDGRTNAYMOE
          QPGDAPRPMSTIGNSASHPPVSCMITOTTERTHETIRALHRSPLYSQGLEAGPR
          YCPSIDKVRFAEKNSHQIIVEPEGLNIDYIPNGISTSLPFDVQLVLSINGFEG
          AHITRGAIVIEDEFDEPGLKASLEKALAGLEFGQINGTGTGEAAAGLGLGNA
          ARVRELCSWTPEDQAYLVLDLITHTGNEPMTSRAYEVLQIRENDAARLT
          AIGRDLGLIDDAHWAHFNKOEVAARECERLSAFWATGPNALGREVKETLVTISRET
          NIIDLMKREPLDYAALMRPYSIPGVGDVADVAEOEISVKYAGYVQSEETRORH
          ENATFPLEDYAAYVRGISTEVLQKHIDPQIVGAQRIPMTPAISILLVHLERM
          RNVA"
          complement(4348. .4560)
          /gene="XF2107"
          complement(4348. .4560)
          /gene="XF2107"
          /note="hypothetical protein; identified by sequence
          similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84906.1"
          /db_xref="GI:9107235"
          /translation="MTLLMCVHTHPLIDELVAMRAGPVIRSMYNAIKODSSGITE
          RLVPWFSWGRASKIDATTAAILASV"
          4928. .5001
          /gene="trna-Cys"
          /note="found by tRNAscan"
          /product="trna-Cys"
          4928. .5001
          /gene="trna-Cys"
          5207. .5392
          /gene="XF2108"
          5207. .5392
          /gene="XF2108"
          /note="hypothetical protein; identified by sequence
          similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84907.1"
          /db_xref="GI:9107236"
          /translation="MOADGTIICLDISTKRWAVAFLSGNPLSTEDKSDVFGAALR
          NSAKMDTQRLAPIKTA"
          complement(5566. .5892)
          /gene="XF2109"
          complement(5566. .5892)

```

```

gene      /gene="XF2109"
CDS       /note="hypothetical protein; identified by sequence
          similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84908.1"
          /db_xref="GI:9107237"
          /translation="MTGAPVRSSFFMKRFKORLGRVRCRCPLPMSNRVCLSRVI
          ALRVVAASCWVWENKKRPVPAAVERCRAAISRDPAAYRMTWGHYDMLKQLQAA
          SELHAA"
          complement(6308. .6646)
          /gene="XF2110"
          complement(6308. .6646)
          /note="hypothetical protein; identified by sequence

Query Match      64.0%; Score 16; DB 1; Length 13856;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 gatctgctgttgaaag 17
Db      2652 GATCTGCTGTTGAAG 2667

RESULT 26
CBRG01D9/c      CBRG01D9      40031 bp      DNA      INV      02-MAY-1996
LOCUS           Caenorhabditis briggsae cosmid G01D9.
DEFINITION
ACCESSION       U56248
VERSION         056248.1
KEYWORDS        GI:1293789
SOURCE          Caenorhabditis briggsae strain-Gujarat G16.
ORGANISM        Caenorhabditis briggsae
                Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
                Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
                1 (bases 1 to 40031)
                Wu, X and Le, T.T.
REFERENCE       The sequence of C. briggsae cosmid G01D9
                Unpublished (1996)
JOURNAL         2 (bases 1 to 40031)
AUTHORS         Washington University Genome Sequencing Center.
TITLE           The C. briggsae Genome Sequencing Project
JOURNAL         Unpublished (1996)
REFERENCE       3 (bases 1 to 40031)
AUTHORS         Waterston, R.
TITLE           Direct Submission
JOURNAL         Submitted (23-APR-1996) Robert Waterston
COMMENT         Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO 63110, USA
                e-mail: mmarrata@watson.wustl.edu
                NEIGHBORING COSMID INFORMATION:
                The neighboring cosmids, chromosome and orientation of G01D9 are
                unknown.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFINDER(P. Green and L. Hillier, ms in preparation).

FEATURES
source
1. .40031
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G01D9"
10. .14844
/gene="G01D9.5"
join(10. .816,859. .1074,1134. .2667,2729. .2930,2975. .13571,
13615. .13860,13905. .14078,14121. .14306,14353. .14400,
14449. .14558,14609. .14693,14746. .14844)
/gene="G01D9.5"

```

```

/note="Similar to polynucleotide synthase."
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98699.1"
/db_xref="GI:1293790"
/translation="MDSLIVDFVNLNKKYEPPEBITASDVDPYVYDELADHTVRK
KYSFELHPASGFTMGVEMPIONEPVEBOLITQSYILKNSPGFKLDVLLTWSGLQ
LNSPSSNPLTALINDIGQDDELKRLMAEKONILINFDLQDSSIEITLYMLLNKVT
LSKSKIKCNFVSGEFLSGNSIARAFMKTISEKKYPLISEMDOKIOOVSPDSSSI
PGMMLITGELTSGVFEIGKFLGNSAENKLTIRROPADVYKETKMTAVHTIAD
VNNKKKLIIEELTAVNDVNGIISAGVILKDSKTIRQTKDSPEVVRPADVAKLEEI
EKHSYKIEHTIMSSFTAACNNEGOLANTGANALDEHHVORREGSGGALQGNW
IDTCMATNQVRYKFTIADGFLQGHNKDAKLYLKAECIEKPEIYVANAIVAMPYIILNRK
DLPRLDIINGLIFSTVITLBNNGNQSFPESDNFESEINVADEADADVELLEKYS
ELVYCAPSKLSKNKINIMDQMSRLVEFLNANTLFEKIKLNLADAYNSLFTLANK
IIPDLNTEAESPMGODEETLVDKRETKSTIDCPKGVNVPKDKENLAKSEA
VKCLVSGKTLPAAGNPAIPVIGTSTSDVSKTKAKETVOVTCQKRSKCVMLTGQSS
OYPLMGROLVENEVFRATLQSLACKCBYKLOGEYSLWEILLFTEHYLLQTLHMOP
IMPSRTIAQMLHSFGVLPDYTLGHSVGLAAGVLAGIMTLEDGLRLYERGRAMENP
IAGGLALVQREIADVLRKPKSVATINSKQVYFAGTPELDSALAFVGGKQKG
TVNORYPHSSILIDETHLVALRKLADIREFOGATRLVSNVTGOIINTESEAVITIKH
TISAYKFCVCEVTLRNDGVAVWIDAGPAALIAFPKRTIIVPAELSKRHTVOTCKEKS
DVELTVOCSLELEOSGFLLMPKVVYGCDDTDAPLAHPVQSTIQHHEELLIEGHR
LNGKIIIVAGYQMPKMDLVKSKASGLDVLKKNKFPVWPVIEDCKAFLLRNMSDMSI
ELVYNSVTVYCSAKVEMKSGALETOVIGENKFPDVEHEVETIYLRNGLOYSGEFRMD
SKRSORCFSSLRKPSFPAMPLDSAMHSITASVPRDPSYFLPVAGOVYTLKOEHP
KLNNAOAVITVITSEDKFIOVVNVVYAGDOILICEVARNMTIYVLRKLVAPENPTPTTA
KAKNASEIIVGFIDTLPLYNOLSTNENMSFLKTRNXPOLQTNATKEERVALLTD
LRVMPDEYFGIRSEAGYLDQORLLITSVAKLSDSISLSPKTVFVIGCSTYERS
OIVAHNODENPRABWGGSTNSALAGRIAHMLKIKGVLIVDVACSSSPHALSACCT
TISGQCECAIVGTVALVHEMTSVLQAOAMTVQDKAPVDANGYKRSBAAVCTLLI
STYGLKSMASIVMAVAGHNGNSSLFTPNQISOLEIKAKATSVRTDIICEVHFGT
KLGDPLELSAISKLSTDGSSSSKSNIGHAEGSGLVSLCTILMSLOTYKMAOHL
KCPTEVITKGAJHCQFEGEDLADBEHSAFLINNGFSGNSGLYTRKQNDPELHESK
DVEYPMILSAHTIKSLERYISVPRPEFASSASIRDMLSILQKAVHAKOPLIYSK
RKLVSODPDDERRMNPITORCAVYKEGTVEKATHFQVROLPPVFNLSLHKLNF
HDEIGQOCVPLREITFEKVLVEKAENNEPMPNVAOVGRITLIPKITEMEYSCFNP
ITVPHLSNSIEEELKLELMSLVSPDONSIVIVACFNKGTSHTEMVYGRJASLEP
ILKYVPSIDVYDHLKEKFNQYDFEESIEFYNSGSRVRLKRLPTPKMKEITKCLIS
GOTGIGRAVANEIKPKKSVITIRKEPEKEETKSYLSIDVADPISHKPHVHFKGT
VKNASHLNIKKDILEMYLTFKOGAKNINCCSKLCKFTYSSSIACILGSGFGSNGT
ANGVNSFLNLSRNTKTIHNGPKKECYMLAKSCDDIKROIEMNGVPLISNDALSV
FSTDLINKHOVYFDPGNEFETVSROPHLOMLSEITLAKSTAKQKQSPDESIFVYV
GTRMDNKKHILPEMDLIGDSLCEMNLKHTLNLNLNELANSIEIENATYELKLYVAS
MLKSKSTESKSPKSDNITNRENDENRVAIVAGSAEENIEEEMENLIDJVSYTG
KENSILKSPGFDNKFENITDEDAKMLDPQVRFIOMSTYALNENSGYVRQDLKRAY
FAGSEPDYGPRESODAMRKLFPYMNANNLTAASYACDILKGETVSVASCTITLVA
RVANAKTIOSSGVYALVGTASIAESECISGNSDEKTIIESKSGICRPFDRNSEGIV
RSGVGGLVLYKRYSEAKKDNDNHFVILKDAISNDGISRASFPAMPACQOQKMEVL
EKLQEBENRLEVEYCHAGTTLGDTIEMNALQTAAYCFKNLQIGSKANITGHAYAS
GLASLICKAKMLQKAVPQVNFSEPRDSFGTFPNVSTOKSFIONDSFLAIDSPGIG
TNVHMIIERLIGIREKNDDEVENVILPISANSKASLVEAIKTMEOVLKKSNIKLSST
LLHHRNOMNERAFITASKIIGKVOIKSFKIKTSMKSPKIALPFAPOGTOPANILPHE
YVHNFDEKSLERHCKTAEHLGVSNNLEILPENTYNEBDLIDITLISQSLVQCMAL
PECLIAVIRPOLIGHSVGEYAAAVISQCSIESEMELLERSNLIISOTOKAMAMV
NYREBLDINSVYSKAIIDPITKCIYVPEALIDELEKTFKCOVKRYTNVILDFVNNIT
QVSOKEKNLCLSPSEPPRIPIISSITGQRIKRPFAEYCRKHLITVAKNLHFGHMKML
KSGVDVAYEIGTPGVSLNLGKKNSKIVIPTCGTGCKHPKASISQCTGLTACCGDILR
PLPEKLIKIDENAGVCEPDEYVQGHKKTPEVLYKFNYSKSKKINSKINSSTTALASDFQ
IYDNIGSTQTSKILYVHLHGSGDSHONFVEYKPLIKQCYETCIEITSSDASVHII
IGLIRHOIYSPRTKIKFVENFEMLELKAVNSAITSGLYVRSVSSKISSIDIRLKAKVE
VLSNLKHFVYLVNCVGGESSKINKTFEBOEPLITTKYNSNLAKLOEKYKTES
TKDMLSNLSNNEELDNALKITALLETEGTLVLEDDPAOTAKRSISYHRDEIGEV
ENBORECLVTSIRINIKETLIGOSIGTIDISNFSIGGDSLSALQVWVOYKTKCTIE
TSTRNYVFEIYVATIDISFCWNSLKYSIISLIVAYQTSYRSVFNHSPPILOKUSLSEF
HNLDWDMDLDAISOELHYSFETGQTPPLRVIAEDSDNMOIHVFNOMHILITDQMSK
TVLADTISKLYSFKLTVPLKSNQSVAKLAIQOCVENLOKHSIENHILITDQMSK
IIPYDAVKEENHKKRIKRIPIRINISKIYDMKSKSTTHNCLAVPECYKASFTGQI
DLMAFAPNSRNSENKDIYGVPMNNALEKTSISEVEKPEKLTIVSVSLDKCRCSN
VPRHTIYVEONRKINAIISTYFNRQKLDYPSVSIPOATCEIKHLISLNNAFSPSTIDET

```

```

trna
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS

```

```

PTGPIILSIYDSSKYPFESTVNLPSFELPLKAKKAFSFKAAONKKIIVHRRADPSPFLQ
IDLITSMKLPESDESPALLSASTIMYSELALIEKVAACIKQKQVIOEAAVASREDIIT
GVYSSKTPISILACSLIGASVAVPDPSEBOLYLVKSVINIVEDSHVFEVLNR
SEKTRQYAGIYITFSGSTGVKPFVMAEGSVSSFLSATKOCMPREKRVYLDVSVH
VFDVSYNIFGPMNLNGCCLISADHSTTTIDPKLOKQVAFPLAANPNGETSETARLRT
VENTLITGGEFVSADALOCALKMPKMRILIOITGPETVWSLJINCRJISEENGANL
IGRIDOVAKCARVDLELEKELLGCLGLQVLAIVSNOMLAFELAGTSAFQSDI
IAKLOKRTLPHVYVOIDQMPUNSSGKLEKSTLLQAFESVRSYKREVPKRNLEEK
IYASGDVYLGRTTEITDKESDVGNLSLFAIOLAHLPTELKEDYKARHEIIOSSNGCL
SKLASVYTPKPSKPNVTTKLEVPNSKFSIYVHAIGTITPYYSFLOIFRPNISL
YGLEPDKYPCGDILREIAPHYQELIAAAGOKOIFWQHSKGGIIMSREIVAEKTIWY
OJPEVWLPESWYLRTEINDELINIKOFITVYFSGLPDSHRIDRAIKLAOLREKTSY
SDTKLIRSKQIGDAARKKAARVDLNEILSRMTONGFDELSTQPIETVLLIDGDHES
CLAKENLKKVXNDPILAPKPYDQA"
1813. .19760
/gene="G01D9.4"
/join(1813. .18387,18490. .18569,18615. .18737,18791. .18886,
19060. .19237,19285. .19395,19444. .19624,19684. .19760)
/gene="G01D9.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98700.1"
/db_xref="GI:1293791"
/translation="MTITFTETPTDKKNAEGYKEDNPVKVYVCEADEEKHISNGVWP
MEGWFDPKPKERSSSEGLSIIKLSPSPLARPPSPDPYEAVANARRRKCTKLE
TLIAFESVILVYSEFLYXOGLISVAYOKYGHGNAATFNRKQTOLEONVONIND
SEKIDVPEFGSNRPALFLHDFKONLAIYDVGNRFCVLDLRTKTRSPRLIEMLR
NIDIMAPVAAVQOTRMVREYVNGVDELTKADISTFNTSILSRICARQYTKLRKQZAO
TISRRAACMGALSPASMAQSDVLEETSF"
complement(19138. .19236)
/note="Leu: codon recognized: CTA"
20621. .21971
/gene="G01D9.3"
/join(20621. .20772,21464. .21614,21664. .21783,21834. .21971)
/gene="G01D9.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98701.1"
/db_xref="GI:1293792"
/translation="MKIFNLSPVISTYTPKPCCRDHLGAAMCSKIMHONTLPKARC
NSDAFPLIOCCSCNDGCMAYDILARSISEKFPDYGRFCDLRSLPPLSNTP
HNSQCGENPQATFRICRKSQCYCNFSVQYTLDNAADACRVQVPAEKKRMWKYFV
SSTITTSFSPSEVINSTYQEWNNH"
complement(22269. .22799)
/gene="G01D9.6"
complement(join(22269. .22371,22415. .22541,22588. .22680,
22730. .22799))
/gene="G01D9.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98702.1"
/db_xref="GI:1293793"
/translation="MIPRFLKVCILISLIGPGYVDASTYSRRPSPERWSIIR
PRMFOPEKTIQOTLOKSTOLDNRRALPCCRKDESGSICKNLRTDILKFOKQTEP
DLSLVCNCSAGISYKRRKQOVNL"
32520. .36916
/gene="G01D9.2"
/join(32520. .32678,33205. .33505,33554. .33662,33836. .33936,
33983. .34088,34698. .34775,34824. .35288,35332. .35502,
35548. .35847,36442. .36637,36699. .36916)
/gene="G01D9.2"
/note="Similar to receptor tyrosine kinase."
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98703.1"
/db_xref="GI:1293794"
/translation="MONGVDVDAQITASSSFDKQSGVQONARLHSELASGACPKPOL
NKSSTFEFLVNDLNFILTSVETQRYNGNGREBASHYMDIRPQSGOWIRYLYNRSG
HYMDGNPDTTVPTRALDPPIVASRIIVSSKTRVCARAIKRAHKGEGVLYYST
VBDGSLDILDFKDMNFENSOYTESGTLGGLLIDGVYQTSDFPKRONNNSWICW
NBDTDGKVTILFEFEELIHNFIDVLAFFGNRIGIDVYISQDKTPTPLSQSISSR
QTLNNTSRDYDRVRLNHRKAKVATISIKFSDDMFLTEVHFTSGITKHTVSONKINY

```

```

LAANLTMTKVPPEPESATTOKLSVIGIIFLLIFVGSIVCSVCLKRRKNSKSDNS
VKKDLITTHMGKPCSCHEVPSNGKLSNGHYVANDILITRSOKSTLSSVSSSSSC
VTPPTNDNFPPPEPPKPEPHTYSOPSPYNSNGSYKSKVPALKKYPSTALLTG
KAIGBCKRTMKECIEGGLCAKRSKSEECVCHSTRAMGEISCLKCGHNPVEL
TGVDSYVLLLEPEVGGCIRNFWLWASETSLTEFLARCKDVFSAWTLYESIRVHGH
FTPNMILLDGEFHAKICSPRGSHAOQLRYGAPSIYNNETPHKSDMAVASTYVEA
YQORRPEELTNEQIVDMARALDHOPHAVRGKFEELSVIMFSAICSIDSNSVL"
38402..39782
/gene="G01D9.1"

Query Match      64.0%; Score 16; DB 5; Length 40031;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      6  tgcgttttaagcgca 21
          |||||
Db      7934 tgcgttttaagcgca 7919

RESULT 27
LOCUS      AP002897 156425 bp      DNA      PLN      08-MAR-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0686E09.
ACCESSION  AP002897
VERSION     AP002897.2 GI:13365563
KEYWORDS
SOURCE
ORGANISM   Oryza sativa (cultivar:Nipponbare) DNA, clone:P0686E09.
            Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoae;
            Oryza.
REFERENCE  1 (sites)
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0686E09
JOURNAL    Published Only in Database (2000) In press
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Direct Submission
JOURNAL    Submitted (19-OCT-2000) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsesaki@abr.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT    On Mar 16, 2001 this sequence version replaced gi:10944300.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RCP. Protein homologues of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DBI accession no. and RCP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            This sequence of P0686E09 clone has an overlap with P0410E01
            (DBJ:AP002866 ) clone at the position 1 to 1233 of 5' end.
            Detailed information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgrp.dna.affrc.go.jp/Genomeseq.html.
            Location/Qualifiers
                1..156425
                /organism="Oryza sativa"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /chromosome="1"

gene
CDS
    join(4841..5341,5904..6008,6478..6591,7059..7241,
    7450..7578,7719..7781,8016..8111,9084..9173,9591..9707,
    9900..10022,10215..10307,10717..10878,10992..11114,
    11186..12178)
    /gene="P0686E09.1"
    join(4841..5341,5904..6008,6478..6591,7059..7241,
    7450..7578,7719..7781,8016..8111,9084..9173,9591..9707,
    9900..10022,10215..10307,10717..10878,10992..11114,
    11186..12178)
    /gene="P0686E09.1"
    /note="hypothetical protein
    similar to Arabidopsis thaliana chromosome 1, F26A9.21"
    /codon_start=1
    /protein_id="BAB39109.1"
    /db_xref="GI:13365564"
    /translation="MALNMTLTQALAKASVTEKVTVTTVQVETGPPRLDPELLEQ
    AGSCGPGIARVITARREGAPSPAPPVSVWVLDRLALAEARAGLTAKAEDAFD
    LARADARLRLRHPGVLAHVQALDETKAMAMATEPVFASVNALGCLDNGKVEKE
    LKGVREMGLEKVKHGLQVSELTDELHNNAHLARAI SEQTVF TLINGSKLGF
    GFALSYDQATGSLSTQGFHYSDYVEDTALPIQPSLNTAPRLVRSQKVSACDI
    FREGCALYHLVARRPLDCHNNKMYNLSITTSRFSNTPADIVYDQRLSDVDA
    SRSAMAFGTGSSFFRNDTRLRLRLFDHLLDERDMQKTEFLKALSDMKKDFDSRV
    LRYKVQVLPRLCAELRMVWQPMILPMVLTIASQDQKDFELSTPALVPVTSASG
    ETLLLVKHADLLIHKELSHVLMPLVRAYDDTDRLDEVLRTVPLSRDLDVQVQ
    LTKOSVLPVHGLAKTTVAACVCIHVLVLRVNALRLGDLVPSLDKEGLEIELQTL
    RCPAVNHTAPRLMCTLGIVANAIEFKQCGVEFAEYVYIPLPFLTAHNOVQFAKY
    ILFVKDITRKIEERKGVTVTDNGTEKVSPLANGIHSEPLSRGQTEQIOTAKGA
    AMDDWQPTTKRATPSQSSSDNSAMKQSPDFPSTQKPSALPDFSTQKPSA
    IPQVTPATITSPASLPSQSLAPSAGMGPSPVDPIDEMPVRKNTSDPNALVSE
    QSKSGLSGDLAIDPFADMPKASGASIIILAVGHLSYTSQSIISGLSTNIGFSNS
    GILGOMKTNGMSAKPNTNTLMAVNTGTVYLNQGNALGFGNPICGSRSTGSLSSSTT
    QGPRRLAPPSASVGRGRGRCQOSALSRSRPHSVSVSEQPIIDLL"
    complement(join(13392..13540,13940..14123))
    /gene="P0686E09.2"
    complement(join(13392..13540,13940..14123))
    /gene="P0686E09.2"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39110.1"
    /db_xref="GI:13365565"
    /translation="MHPISRFAPVYLAVPFRSGYPCRLAAYPAAPCPRLDPEYH
    NCVAVAPLQFGVAPARCIMLDGATVTEFEGQPASSQLAKSLCVSHSDVAMEFH
    VHRVTDPP"
    join(28972..29115,29117..29179)
    /gene="P0686E09.3"
    join(28972..29115,29117..29179)
    /gene="P0686E09.3"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39111.1"
    /db_xref="GI:13365566"
    /translation="MHAASASARTACSAARSTSLERWGAHGHAGVTTIDGMR
    RAIPHRIQSMRARRAGRGREG"
    complement(join(29740..29837,30219..30380,32426..32546))
    /gene="P0686E09.4"
    complement(join(29740..29837,30219..30380,32426..32546))
    /gene="P0686E09.4"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39112.1"
    /db_xref="GI:13365567"
    /translation="MFGAISGLSPREVKVNLNLEVVLDGHSRTQGAHLDONLOR
    KRGDISTENALNAHKCPPLAVLFTQLQMTQASAPVSSRPAQLKRRIDPTECV
    GQGEQRTDKLKLKASRCDAATR"
    join(33917..34076,35467..35816)
    /gene="P0686E09.5"
    join(33917..34076,35467..35816)
    /gene="P0686E09.5"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB39113.1"
    /db_xref="GI:13365568"

```

```

gene
/translation="MWTCLAGAGVLPGRATHVLLTLRSGEROFTLLPALLFPCPL
GCMGLIAVAVLYKDNIEFOEAVRKECVPCVGNKMAKSTVEVGAERSTWPLEIY
GVLLIPSLISVFGKQLLMFLGYACVSVREKAKGLDRACHPFLSLINMNDLGLKC
NMHPFLKIS"
join(37532..38415,38988..39218,39280..39471,39552..39890,
40001..40207,40381..40515,40601..40765)
/gene="P0686E09.6"
join(37532..38415,39005..39206,38988..39218,39280..39471,
39552..39890,40001..40207,40381..40515,40601..40765)
/feature="hypothetical protein"
similar to Oryza sativa chromosome 1, P0665D10.16"
/codon_start=1
/protein_id="BAB39114.1"
/db_xref="GI:13365369"
/translation="MSEADRLKRAAALALDGGGAGRDPAKADAVADLDPVSP
LRASVATSSSSSSGSAKSPAPNAGALALAGRSGETTASTPTPLPGRRCSS
GPLIFSGSGSGGGGDRGTASTPTNALLPGNIGSPGRVAVAAAPPPSRD
VLGGTGYGHSIMRGSGMAPARSSIDSSFTGHA PRSPATPRASSASGSLDQY
TRICNEYKKKKKHAALRHYDRAVALPESACGNNAALAGRLADLRCEERY
RLDPANGRAHRLAGLRLGMLSKARHLTQGLHOSDSEWEKLOEVMHOGSEI
DARKVGWKSALREDAALAGADSSRLTLGLMIISKARHLTQGLHOSDSE
WEKLOEVMHOGSEIDARKVQDMKSALREDAALAGADSSRLTMSLLIRSFALR
LHLEFADSTLAKLDSVLLYPMGANPSGMLAESVYSVRAQVDALRCRCPEDA
VEADNARFIDPNAEYGMILNNYKLVAKAQNELIYKAKFSADATVASEGICYP
SNPLYCNBAACWCKLEMRKAVDDCMALRIQPNYTKALIRASVYAKYERADCYR
DYEVLEKPADPEVALEAFHAYALKTGTGEDVSNKREGVEVMTSVQLRAIIS
PGVLEORAISSGVYFMKIMNOCTLLTPVSNSLSECPLELTVQVNDSPV
AKAENRIVLPFKIYKGVKVKEMICPSLIVRSHVSVSS"
complement(join(47016..47264,47291..47363,48616..48748,
48784..48832,49178..49324,49569..49684,50921..50967,
51070..51107,51434..51652,51763..51861,51888..51986,
52059..52112,52817..52933,53291..53532,54642..54895,
55164..55345))
/gene="P0686E09.7"
complement(join(47016..47264,47291..47363,48616..48748,
48784..48832,49178..49324,49569..49684,50921..50967,
51070..51107,51434..51652,51763..51861,51888..51986,
52059..52112,52817..52933,53291..53532,54642..54895,
55164..55345))
/feature="hypothetical protein"
similar to Arabidopsis thaliana chromosome 3, MBK21.5"
/codon_start=1
/protein_id="BAB39115.1"
/db_xref="GI:13365570"
/translation="MSDRVSQRENTMDMGYARENKAMDGLAVVEGARESKCA
LGGARVHILIGRVSQASRQPSAKTLASASGASMSKDPYPHHGGAAGEGA
AAGTSTEVLDLARTGRPLANPSSAADLDAAGARSMCVLTHQPMGSHSTVE
QIEALYSNTMTKPRLESSLPYQPRGKDCAFYMTKTCVFGSGCKRDHOSWPE
GGIPNMKEHLFEALDNLTLKGYIGNSCOHHFVYLTALGVQPRAAVEDSYED
EGEDPCPEFMKTGCKGSKCKENHPEKYNALASGNTINDTGMSSOPLRSPQCSF
MKLOKHLIADSSILPEYAKTGCKCKRACNHPDKDEITSSONPEASATVGHEDPI
GSAADSVSAKMPQVAAOEFENKGLPMRGEVDCPRYMGSCCKGCKGCKRERHRL
VLRNPLPGOTIILPTPSMLINSANPMQGFDRHAAMPYMTKTCRCKFRDCKFHP
IDRSAPDSAMEPAESVOLTLTAGLPRREDVAVCAVMTGCKGCKGCKGVEFEL
RRETNILQGRKISMPYSFLAYELNMKPHYLGSEFLYRRLSELESRKAAALISA
OPVAYVGHRLAFOVAPLHPTASGREPIITRGSIALIAMSFRRAAGGDAVRLGCLPV
GCAMPARSAASGDHPLCCAGY"
join(57612..57759,58207..58289,58617..58775,59266..59373,
59488..59548,59653..59918,60187..60219)
/gene="P0686E09.8"
join(57612..57759,58207..58289,58617..58775,59266..59373,
59488..59548,59653..59918,60187..60219)
/feature="contains ESTs D40230(S2054),D23027(C2064)
hypothetical protein"
similar to Arabidopsis thaliana chromosome 1, F19K6.5"
/codon_start=1
/protein_id="BAB39116.1"
/db_xref="GI:13365571"
/translation="MVGMAAGARAVLRLLSAAEAAKODGRVFAASYSSSGSVN
APRIGQYANILRAQAFASRGVALNRIHOLIRNAGISTRNLLAADDAMVPVSSPLTPR
LPGGEOTDKKGAIVKRLKVOAIKKDKIKQSPKKVNLVAKLVGRKRVEDALLQDQYTKR

```

AAVYHSARANAHHGLDPPKLIYEAFVKGGLYLKRLSYHANGCGVWPRCRL
TVVVRATAEERAKIAKLIRSVYKILTRERKOLMHRILEVSPRARRRKEBACAAAL

Query Match 64.0%; Score 16; DB 13; Length 156425;
Best Local Similarity 100.0%; Pred No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 52932 ATCGCTGTGTAACG 52917

RESULT 28
AC016697/c
LOCUS

DEFINITION
Homo sapiens chromosome UNK clone RP11-140C4, WORKING DRAFT
SEQUENCE, 2 unordered pieces.

AC016697.7 GI:13431103
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFN.

REFERENCE
AUTHORS
Waterston,R.H.
The sequence of Homo sapiens clone

Unpublished
2 (bases 1 to 164227)
Waterston,R.H.
Direct Submission

Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

On Mar 22, 2001 this sequence version replaced gi:11875313.
COMMENT

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0140C04

Sequencing vector: M13; 478
Sequencing vector: plasmid; 298
Chemistry: Dye primer ET; 47% of reads
Chemistry: Dye-terminator Big Dye; 28% of reads

Assembly program: Phrap; version 0.990319
Consensus quality: 163129 bases at least Q40
Consensus quality: 163369 bases at least Q30
Consensus quality: 163445 bases at least Q20

Insert size: 16900; agarose-fp
Insert size: 163485; sum-of-contigs
Insert size: 6.69 in Q20 bases; agarose-fp
Insert size: 7.00 in Q20 bases; sum-of-contigs

NOTE: This is a "working draft" sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 642: contig of 642 bp in length
643 742: gap of unknown length
743 164227: contig of 163485 bp in length.

Location/Qualifiers
1..164227
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-140C4"

FEATURES
source

```

misc_feature      1..642
                  /note="assembly_name:Contig20"
misc_feature      743..164227
                  /note="assembly_name:Contig21"
BASE COUNT      46978 a 33168 c 34928 g 49053 t      100 others
ORIGIN
Query Match      64.0%; Score 16; DB 64; Length 164227;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 gcctgttgagcgagc 22
        |||
Db      40943 GCTGTTTGAAGCGCAG 40928

RESULT 29
AC023914/c      168177 bp      DNA      HTG      20-SEP-2000
LOCUS
DEFINITION      Homo sapiens clone RP11-44K20, WORKING DRAFT SEQUENCE, 22 unordered
                  pieces.
ACCESSION      AC023914
VERSION      AC023914.3 GI:10198362
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 168177)
Birren,B., Linton,L. Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-44K20
Unpublished
2 (bases 1 to 168177)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepeel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fensterhor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Kleland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuck,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trifilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced by:9102875.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1087
Center clone name: 44_K-20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The line order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1263: contig of 1263 bp in length
1264 1363: gap of 100 bp
1364 2399: contig of 1036 bp in length
2400 2499: gap of 100 bp
2500 3713: contig of 1214 bp in length
3714 3813: gap of 100 bp
3814 27125: contig of 23312 bp in length
27126 27225: gap of 100 bp
27226 28854: contig of 1629 bp in length
28855 28954: gap of 100 bp
28955 31625: contig of 2671 bp in length
31626 31725: gap of 100 bp
31726 34084: contig of 2359 bp in length
34085 34184: gap of 100 bp
34185 36582: contig of 2398 bp in length
36583 36682: gap of 100 bp
36683 39629: contig of 2947 bp in length
39630 39729: gap of 100 bp
39730 43900: contig of 4171 bp in length
43901 44000: gap of 100 bp
44001 46837: contig of 2837 bp in length
46838 46937: gap of 100 bp
46938 52599: contig of 5662 bp in length
52600 52699: gap of 100 bp
52700 60235: contig of 7536 bp in length
60236 60335: gap of 100 bp
60336 67600: contig of 7265 bp in length
67601 67700: gap of 100 bp
67701 76219: contig of 8519 bp in length
76220 76319: gap of 100 bp
76320 85003: contig of 8684 bp in length
85004 85103: gap of 100 bp
85104 95875: contig of 10772 bp in length
95876 95975: gap of 100 bp
95976 108080: contig of 12105 bp in length
108081 108180: gap of 100 bp
108181 126996: contig of 18816 bp in length
126997 127096: gap of 100 bp
127097 145024: contig of 17928 bp in length
145025 145124: gap of 100 bp
145125 166554: contig of 21330 bp in length
166555 166654: gap of 100 bp
166655 168177: contig of 1523 bp in length.
Location/Qualifiers
1..168177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-44K20"
/clone_lib="RPCT-11 Human Male BAC"
1..1263
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
1364..2399
/note="assembly_fragment"
2500..3713

```

```

misc_feature      /note="assembly-fragment"
                  3814..27125
misc_feature      /note="assembly-fragment"
                  27226..28854
misc_feature      /note="assembly-fragment"
                  28955..31625
misc_feature      /note="assembly-fragment"
                  31726..34084
misc_feature      /note="assembly-fragment"
                  34185..36582
misc_feature      /note="assembly-fragment"
                  36683..39629
misc_feature      /note="assembly-fragment"
                  39730..43900
misc_feature      /note="assembly-fragment"
                  44001..46837
misc_feature      /note="assembly-fragment"
                  46938..52599
misc_feature      /note="assembly-fragment"
                  52700..60235
misc_feature      /note="assembly-fragment"
                  60336..67600
misc_feature      /note="assembly-fragment"
                  67701..76219
misc_feature      /note="assembly-fragment"
                  76320..85003
misc_feature      /note="assembly-fragment"
                  85104..95875
misc_feature      /note="assembly-fragment"
                  95976..108080
misc_feature      /note="assembly-fragment"
                  108181..126996
misc_feature      /note="assembly-fragment"
                  127097..145024
misc_feature      /note="assembly-fragment"
                  145125..166554
misc_feature      /note="assembly-fragment"
                  166655..168177
misc_feature      /note="assembly-fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      41092 a 41268 c 42342 g 41368 t 2107 others
ORIGIN

```

```

Query Match      64.0%: Score 16; DB 68; Length 168177;
Best Local Similarity 100.0%: Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atctgcgtttgaagc 18
          |||
Db      81387 ATCTGCTGTTGAAGC 81372

```

```

RESULT 30
LOCUS      AC090534      188199 bp      DNA      11-APR-2001
DEFINITION Mus musculus chromosome 18 clone RP23-233P17 strain C57BL/6/J,
ACCESSION AC090534
VERSION    AC090534.5 GI:13591831
KEYWORDS   HTG; HTGS_PHASEL; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 188199)
            Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS    Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,
            Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
            Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
            High Throughput Mouse Sequencing
            Unpublished
            2 (bases 1 to 188199)

```

AUTHORS TITLE JOURNAL COMMENT

Grills, G., Li, L., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Apr 11, 2001 this sequence version replaced gi:13560379.

-----Genome Center
Center: Albert Einstein College of Medicine
Center Code: AECOM
Web site:
http://sequence.aecom.yu.edu/cgi-
bin/ws_exe/mouseDB/mouseSeq/mouseseqtable.htm
Contact: htgs@sequence.aecom.yu.edu

-----Summary Statistics
Center project name: AJT
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 179317 at least Q20
*Consensus quality: 176170 at least Q30
*Consensus quality: 170057 at least Q40
*Estimated insert size: agarose-FP - N/A
*Estimated insert size: 187599 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 8.3 x sum-of-contigs - N/A

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 34856: contig of 34856 bp in length
* 34857: gap of unknown length
* 34877: gap of 28243 bp in length
* 63120: gap of unknown length
* 63140: gap of 15426 bp in length
* 78566: gap of unknown length
* 78586: contig of 15950 bp in length
* 94536: gap of unknown length
* 94555: gap of unknown length
* 94556: gap of 18134 bp in length
* 112690: contig of 18134 bp in length
* 112709: gap of unknown length
* 112710: contig of 12193 bp in length
* 124903: gap of unknown length
* 124923: contig of 338 bp in length
* 125261: gap of unknown length
* 125280: gap of 12176 bp in length
* 137456: contig of 12176 bp in length
* 137457: gap of unknown length
* 137477: gap of 8209 bp in length
* 145686: contig of 8209 bp in length
* 145705: gap of unknown length
* 145706: gap of 6746 bp in length
* 152451: contig of 6746 bp in length
* 152472: gap of unknown length
* 152473: contig of 7024 bp in length
* 159496: gap of unknown length
* 159515: gap of unknown length
* 159516: gap of 3347 bp in length
* 162863: gap of unknown length
* 162882: gap of unknown length
* 164493: contig of 3610 bp in length
* 164512: gap of unknown length
* 165513: gap of 3436 bp in length
* 165948: contig of 3436 bp in length
* 169949: gap of unknown length
* 169969: gap of 3640 bp in length
* 173609: gap of unknown length
* 173629: gap of 1840 bp in length
* 175469: gap of unknown length
* 175488: gap of unknown length
* 175489: contig of 1227 bp in length
* 176716: gap of unknown length
* 176735: gap of unknown length
* 176736: contig of 786 bp in length
* 177521: gap of unknown length
* 177541: gap of unknown length
* 177542: contig of 209 bp in length
* 177750: gap of unknown length
* 177751: gap of unknown length

```

```

* 177771 178025: contig of 255 bp in length
* 178026 178045: gap of unknown length
* 178046 178482: contig of 437 bp in length
* 178483 178502: gap of unknown length
* 178503 180037: contig of 1535 bp in length
* 180038 180057: gap of unknown length
* 180058 180950: contig of 893 bp in length
* 180951 180970: gap of unknown length
* 180971 182851: contig of 1881 bp in length
* 182852 182871: gap of unknown length
* 182872 183253: contig of 382 bp in length
* 183254 183273: gap of unknown length
* 183274 183953: contig of 680 bp in length
* 183954 183973: gap of unknown length
* 183974 184181: contig of 208 bp in length
* 184182 184201: gap of unknown length
* 184202 185477: contig of 1276 bp in length
* 185478 185497: gap of unknown length
* 185498 186883: contig of 1386 bp in length
* 186884 186903: gap of unknown length
* 186904 187053: contig of 150 bp in length
* 187054 187073: gap of unknown length
* 187074 188199: contig of 1126 bp in length.

FEATURES
Source
1. 188199
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="18"
/clone="RP23-233F17"
/sex="male"
1. 34856
/misc_feature /note="assembly_name:Contig108"
34877. 63119
/misc_feature /note="assembly_name:Contig107"
63140. 78565
/misc_feature /note="assembly_name:Contig106"
78586. 94535
/misc_feature /note="assembly_name:Contig105"
vector_end:SP6
vector_side:right"
94556. 112689
/misc_feature /note="assembly_name:Contig104"
112710. 124902
/misc_feature /note="assembly_name:Contig103"
vector_end:T7
vector_side:left"
124923. 125260
/misc_feature /note="assembly_name:Contig102"
125281. 137456
/misc_feature /note="assembly_name:Contig101"
137477. 145685
/misc_feature /note="assembly_name:Contig100"
145706. 152451
/misc_feature /note="assembly_name:Contig99"
152472. 159495
/misc_feature /note="assembly_name:Contig98"
159516. 162862
/misc_feature /note="assembly_name:Contig97"
162883. 166492
/misc_feature /note="assembly_name:Contig96"
166513. 169948
/misc_feature /note="assembly_name:Contig95"
169969. 173608
/misc_feature /note="assembly_name:Contig94"
173629. 175468
/misc_feature /note="assembly_name:Contig93"
175489. 176715
/misc_feature /note="assembly_name:Contig92"
176736. 177521
/misc_feature /note="assembly_name:Contig91"
177542. 177750
/misc_feature /note="assembly_name:Contig90"

```

```

misc_feature 177771. 178025
/note="assembly_name:Contig89"
misc_feature 178046. 178482
/note="assembly_name:Contig88"
misc_feature 178503. 180037
/note="assembly_name:Contig87"
misc_feature 180058. 180950
/note="assembly_name:Contig86"
misc_feature 180971. 182851
/note="assembly_name:Contig85"
misc_feature 182872. 183253
/note="assembly_name:Contig84"
misc_feature 183274. 183953
/note="assembly_name:Contig83"
misc_feature 183974. 184181
/note="assembly_name:Contig82"
misc_feature 184202. 185477
/note="assembly_name:Contig81"
misc_feature 185498. 186883
/note="assembly_name:Contig80"
misc_feature 186904. 187053
/note="assembly_name:Contig79"
misc_feature 187074. 188199
/note="assembly_name:Contig78"
BASE COUNT 51323 a 42282 c 41293 g 52561 t 740 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 78; Length 188199;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggaatcgcgtttgaa 16
|||||
Db 108548 GGATCGCTGTTGAA 108563

```

```

RESULT 31
LOCUS AC007104/c DNA 17-JUL-1999
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 16
unordered pieces.
ACCESSION AC007104
VERSION AC007104.4 GI:5523795
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
2 (bases 1 to 200000)
Unpublished
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
REFERENCE 1 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL Submitted (17-MAR-1999) Department of Genetics, Stanford Human
Genome Center 855 California Avenue, Palo Alto, CA 94304, USA
on Jul 17, 1999 this sequence version replaced gi:4757683.
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1369: contig of 1369 bp in length
* 1370 3774: gap of unknown length
* 3775 4878: contig of 1104 bp in length
* 4879 7283: gap of unknown length
* 7284 8555: contig of 1272 bp in length
* 8556 10960: gap of unknown length

```



```

* 10961 12207: contig of 1247 bp in length
* 12208 14612: gap of unknown length
* 14613 16262: contig of 1650 bp in length
* 16263 18667: gap of unknown length
* 18668 20176: contig of 1509 bp in length
* 20177 22581: gap of unknown length
* 22582 24278: contig of 1697 bp in length
* 24279 26684: gap of unknown length
* 26684 28199: contig of 1515 bp in length
* 28199 30604: gap of unknown length
* 30604 32359: contig of 1756 bp in length
* 32360 34764: gap of unknown length
* 34765 39958: contig of 5194 bp in length
* 39959 42363: gap of unknown length
* 42364 57085: contig of 14722 bp in length
* 57086 59490: gap of unknown length
* 59491 74905: contig of 15415 bp in length
* 74906 84273: gap of unknown length
* 84274 86679: gap of unknown length
* 86679 118061: contig of 31383 bp in length
* 118062 120466: gap of unknown length
* 120467 152965: contig of 32499 bp in length
* 152966 155371: gap of unknown length
* 155371 200000: contig of 44630 bp in length.

```

```

FEATURES
  source          Location/Qualifiers
                  1..200000
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="4"

```

```

BASE COUNT      38319 a 42757 c 43367 g 37466 t 38091 others
ORIGIN

```

```

Query Match
Best Local Similarity 64.0%; Score 16; DB 60; Length 200000;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atctgctgttgaag 18
        |||
Db 129264 ATCTGCTGTTGAAGC 129249

```

```

RESULT 32
LOCUS   PEHE7D      306 bp      DNA      VRL      10-FEB-1999
DEFINITION Human papillomavirus type 21 DNA for E7 protein, complete cds.
ACCESSION D50548
VERSION   D50548.1 GI:808883
KEYWORDS  E7 protein.
SOURCE    Human papillomavirus type 21 DNA.
          Human papillomavirus type 21
          Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
          Papillomavirus.
REFERENCE 1 (bases 1 to 306)
AUTHORS  Kiyono,T.
TITLE     Direct Submission
JOURNAL   Submitted (11-MAY-1995) to the DDBJ/EMBL/Genbank databases. Jotnu
          Kiyono, Aichi Cancer Center, Research Institute, Laboratory of
          Viral Oncology; 1-1 Kanokoden, Chikusa-Ku, Nagoya, Aichi 464, Japan
          (E-mail: h44714@nuc.cc.nagoya-u.ac.jp, Tel: 052-762-6111(ex.8838),
          Fax: 052-763-5233)
REFERENCE 2 (bases 1 to 306)
AUTHORS  Adachi,A., Kiyono,T., Ohashi,M. and Ishibashi,M.
TITLE     Detection of human papillomavirus type 47 DNA in malignant lesions
          of epidermodysplasia verruciformis with protocols for precise
          typing of the related HPV DNAs
JOURNAL   Unpublished (1995)
REFERENCE 3 (sites)
AUTHORS  Adachi,A., Kiyono,T., Hayashi,Y., Ohashi,M. and Ishibashi,M.
TITLE     Detection of human papillomavirus (HPV) type 47 DNA in malignant
          lesions from epidermodysplasia verruciformis by protocols for
          precise typing of related HPV DNAs

```

```

JOURNAL      J. Clin. Microbiol. 34 (2), 369-375 (1996)
MEDLINE      96381005
FEATURES
  source      Location/Qualifiers
              1..306
                /organism="Human papillomavirus type 21"
                /db_xref="taxon:31548"

```

```

  gene        1..306
              /gene="E7"
  CDS          1..306
              /gene="E7"
              /note="putative"
              /codon_start=1
              /product="E7 protein"
              /protein_id="BA09116.1"
              /db_xref="GI:808884"
              /translation="MIGKEVTLQDILVELNELQPEVQVDLFCBEELPSQOEETTEL
              PRTSKVVTGCGCKKRLRFVNATRFATRFQNLRFELQLLPCEGRCKNGKS"

```

```

BASE COUNT      88 a 59 c 81 g 78 t
ORIGIN

```

```

Query Match
Best Local Similarity 60.0%; Score 15; DB 59; Length 306;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atctgctgttgaag 17
        |||
Db      236 ATCTGCTGTTGAAG 250

```

```

RESULT 33
LOCUS   PEU244276      625 bp      DNA      PLN      11-MAY-1999
DEFINITION Pringsheimia euphorbiae 5.8S rRNA gene and internal transcribed
          spacers 1 and 2 (ITS1, ITS2), strain CBS 747.71.
ACCESSION AJ244276
VERSION   AJ244276.1 GI:4808321
KEYWORDS  5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer 1;
          internal transcribed spacer 2; ITS1; ITS2.
SOURCE    Pringsheimia euphorbiae.
          Pringsheimia euphorbiae.
          Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
          Chaetothyriomycetes incertae sedis; Dothioraceae; Pringsheimia.
REFERENCE 1 (bases 1 to 625)
AUTHORS  Yurlova,N., Hoog,G.D. and Gerrits van den Ende,A.
TITLE     Taxonomy of Aureobasidium and allied genera
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 625)
AUTHORS  Hoog,G.D.
TITLE     Direct Submission
JOURNAL   Submitted (03-MAY-1999) Hoog G.D., Research, Centraalbureau voor
          Schimmelcultures, P.O.Box 273, 3740 AG Baarn, NETHERLANDS

```

```

FEATURES
  source      Location/Qualifiers
              1..625
                /organism="Pringsheimia euphorbiae"
                /strain="CBS 747.71"
                /db_xref="taxon:93493"
                <1..196
                /note="internal transcribed spacer 1, ITS1"
  rRNA        197..351
              /gene="5.8S rRNA"
              /product="5.8S ribosomal RNA"
              /gene="5.8S rRNA"
              197..351
              /gene="5.8S rRNA"
              352..>625
              /note="internal transcribed spacer 2, ITS2"
  misc_feature 170 a 128 c 135 g 192 t
  BASE COUNT      170 a 128 c 135 g 192 t
ORIGIN

```

```

Query Match
Best Local Similarity 60.0%; Score 15; DB 14; Length 625;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


gene
/organism="Serratia marcescens"
/db_xref="taxon:615"
656..1465
/gene="aac3-Vb"
656..1465
/gene="aac3-Vb"
/codon_start=1
/transl_table=11
/product="acetyltransferase"
/protein_id="AA26548.1"
/db_xref="GI:152812"

BASE COUNT 276 a 516 c 506 g 274 t
ORIGIN

Query Match 60.0%; Score 15; DB 3; Length 1572;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgaagcgag 22
|||||
Db 1397 CTGTTGAGCGCAG 1411

RESULT 37
AF239157/c 1616 bp mRNA ROD 13-MAR-2000
LOCUS Rattus norvegicus DEXRAS1 (Dextrasi) mRNA, complete cds.
DEFINITION AF239157
ACCESSION AF239157.1 GI:7230767
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1616)
Fang, M., Jaffrey, S.R., Sawa, A., Ye, K. and Snyder, S.H.
Dextrasi: A G-protein coupled to neuronal nitric oxide synthase via
CAPON
Unpublished
2 (bases 1 to 1616)
Fang, M., Jaffrey, S.R., Sawa, A., Ye, K. and Snyder, S.H.
Direct Submission
Submitted (24-FEB-2000) Neuroscience, Johns Hopkins University, 725
N. Wolfe St, Baltimore, MD 21205, USA
FEATURES
source
1..1616
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..1616
/gene="Dextrasi"
118..960
/gene="Dextrasi"
/note="Ras-related protein"
/codon_start=1
/product="Dextrasi"
/protein_id="AA26390.1"
/db_xref="GI:7230768"
/translation="MKLAAMTKMKCPDSELSIPAKNCYRMVILGSSKVGKTAIVSRF
LTGRFEDAYPTIEDPFRKFSYIRGEVYQDIDTSGNHPFAMRRLSTITGDVILV
FSLNDRSPFEVORLKOQIIDTKSCIKNTKENVDPVIVICGNKGRDPTREYREOREI
EQVLGDDPQRCAYFEISAKNSSLDDMFALFAMAKLPSEMSDPLHRRKVSVOYCVLH
KKALRNKKLIRAGSGGGGHDGDAFGILAPPARPSVHSDIMYIRKTSVSGQAKKER
CVIS"

BASE COUNT 403 a 436 c 413 g 363 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 94; Length 1616;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttga 15
|||||
Db 481 GGATCTGCTGTTGA 467

RESULT 38
AF009246/c 1623 bp mRNA ROD 13-FEB-1998
LOCUS Mus musculus ras-related protein (DEXRAS1) mRNA, complete cds.
DEFINITION AF009246
ACCESSION AF009246.1 GI:2253712
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1623)
Kempainen, R.J. and Behrend, E.N.
Dexamethasone rapidly induces a novel ras superfamily
member-related gene in Atr-20 cells
J. Biol. Chem. 273 (6), 3129-3131 (1998)
2 (bases 1 to 1623)
Kempainen, R.J. and Behrend, E.N.
Direct Submission
Submitted (18-JUN-1997) Physiology & Pharmacology, Auburn
University College of Veterinary Medicine, 213 Greene Hall, Auburn,
AL 36849, USA
FEATURES
source
1..1623
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="corticothorpe tumor cell line Atr-20"
1..1623
/gene="DEXRAS1"
142..984
/gene="DEXRAS1"
/note="induced by dexamethasone"
/codon_start=1
/product="ras-related protein"
/protein_id="AAC5358.1"
/db_xref="GI:2253713"
/translation="MKLAAMTKMKCPDSELSIPAKNCYRMVILGSSKVGKTAIVSRF
LTGRFEDAYPTIEDPFRKFSYIRGEVYQDIDTSGNHPFAMRRLSTITGDVILV
FSLNDRSPFEVORLKOQIIDTKSCIKNTKENVDPVIVICGNKGRDPTREYREOREI
EQVLGDDPQRCAYFEISAKNSSLDDMFALFAMAKLPSEMSDPLHRRKVSVOYCVLH
KKALRNKKLIRAGSGGGGHDGDAFGILAPPARPSVHSDIMYIRKTSVSGQAKKER
CVIS"

BASE COUNT 412 a 437 c 417 g 357 t
ORIGIN

Query Match 60.0%; Score 15; DB 94; Length 1623;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttga 15
|||||
Db 505 GGATCTGCTGTTGA 491

RESULT 39
AF039572 2795 bp DNA BCT 02-DEC-1998
LOCUS Serratia liquefaciens serrawettlin synthase (swrA) gene, partial
DEFINITION AF039572
ACCESSION AF039572
VERSION AF039572.1 GI:2760927

ORGANISM	Potato mop-top virus
REFERENCE	Viruses: ssRNA positive-strand viruses, no DNA stage; Furovirus.
AUTHORS	1 (bases 1 to 6043)
TITLE	Savenkov, E.I., Sandgren, M. and Valkonen, J.P.
JOURNAL	Complete sequence of RNA 1 and the presence of tRNA-like structures
MEDLINE	In all RNAs of Potato mop-top virus, genus Pomovirus
REFERENCE	J. Gen. Virol. 80 (Pt 10), 2779-2784 (1999)
AUTHORS	2 (bases 1 to 6043)
TITLE	Savenkov, E.I.
JOURNAL	Direct Submission
REFERENCE	Submitted (27-APR-1999) Savenkov E.I., Plant Biology, Slu, box
TITLE	7080, se-750 07, Uppsala, SWEDEN
FEATURES	Location/Qualifiers
SOURCE	1. 6043
	/organism="Potato mop-top virus"
	/vition
	/isolate="Swedish (Sw)"
	/db_xref="taxon:37128"
	/country="Sweden"
	/note="RNA1 RNA; genus Pomovirus"
	115..4026
	/function="replication"
	/note="ORF1; readthrough protein"
	/codon_start=1
	/product="polymerase"
	/protein_id="CA58364.1"
	/db_xref="GI:6066603"
	/translation="MDOSILSKITSDYVNAIILHTSATRVSDDLHNMCAVYDIOID
	STENAKRRKIDVKNRLNEDQIOLALPEPERRVTVSSVHRHTSMAAMKLETDVI
	FTSPKNGVIVYICGNWATHAKNDGGEVHCCPLIDPRDAQRTLTIDNRIDGA
	KVASEKAAVAAAOIKKCDLISENAKKADVANDLNGTFCQNKREDVYDHSHTGLG
	KKEVAAHISYDIDHLYDVDSAMERKRYVTKGFELPSADIIILCKKGLSVYNGY
	IIDBSIKSYFYDDPGNGYVHNLNLSMLVYTKTPVKAAGAAVYVLETEMGDMPFT
	ITDSEARWGVILHDSSTKPLANKRDLYVPLIDIRADLTDELVRELLISREFVNA
	LEVAQKNDQNTVAEGLISTYFASTNNAVYIGSARKTSKRDPKLLPMTITTLAVOE
	LOKKOKRVILGLKSKYKEELTLGILSEVHVHREPOSLVORGAGVAKMQRYSQ
	DLVGLHDVPLYLEINDRIKLGSAIKVNFSPLSFSLDEKSVLVEYERKQISDEL
	VSEKIGLIEGYVYVSGSLPKQKVASDGAOMVSGECLYNTMKCEERPKPR
	FSTVLEEMISEGNCFALNSFGDEVHMFDSLKEACGRAMRKQSAVTFSDAEFDNI
	ENTKEELIEEVEVYKTTGEVQEDOMPIGNLPDVDDPSAQAQCLTESASTSDSDCP
	MEQLVACERAPATKKEFEVYMDSOVNSELEIYRREGHNALISGESSESSQSO
	ELVAAGVPLSKMOMVEDSEVRARQANDHDLSNDYVYAKMPPREPEAGDQPRIK
	AKREFLWYLKCKLVAADKSTLVEIMRDFTGQFHSACETPKNACLTSEKVCENH
	CKRYRHPKSGKSAVAEFTREDWRAKLKLOKMKAAENEMSGSDSDRTPQGG
	TGYLFCDITFLMEIEPLINRLLEISFKRYORRAPRITLVDPGCGASTYVYEKAN
	VNOYVYTGREAEDLREPRKERNATATOLKRPVDSYLLNDTOSANVLHPEEAL
	MAHAKVYFCADDLSARSVYISAOGDSOQIATPINKVESITLTKSLKLEIDNVEKRLYRS
	PLDVASVLTFRKNTGYTSVYTSANPLVNSLTKVPRDGMSTYSIKIGTGYLFGES
	EKEERQYLYTKAEKNNVNTVHESQKTDINVLCKRLKATONEIYIPGRGNSPIMVYGV
	RHRRSIVYVYTKAEKLYFDLAELMSVQEGKLMHLEBEVK"
	118..4023
	/product="polymerase"
	/db_xref="GI:6066603"
	/protein_id="CA58364.1"
	/db_xref="GI:6066603"
	/translation="MDOSILSKITSDYVNAIILHTSATRVSDDLHNMCAVYDIOID
	STENAKRRKIDVKNRLNEDQIOLALPEPERRVTVSSVHRHTSMAAMKLETDVI
	FTSPKNGVIVYICGNWATHAKNDGGEVHCCPLIDPRDAQRTLTIDNRIDGA
	KVASEKAAVAAAOIKKCDLISENAKKADVANDLNGTFCQNKREDVYDHSHTGLG
	KKEVAAHISYDIDHLYDVDSAMERKRYVTKGFELPSADIIILCKKGLSVYNGY
	IIDBSIKSYFYDDPGNGYVHNLNLSMLVYTKTPVKAAGAAVYVLETEMGDMPFT
	ITDSEARWGVILHDSSTKPLANKRDLYVPLIDIRADLTDELVRELLISREFVNA
	LEVAQKNDQNTVAEGLISTYFASTNNAVYIGSARKTSKRDPKLLPMTITTLAVOE
	LOKKOKRVILGLKSKYKEELTLGILSEVHVHREPOSLVORGAGVAKMQRYSQ
	DLVGLHDVPLYLEINDRIKLGSAIKVNFSPLSFSLDEKSVLVEYERKQISDEL
	VSEKIGLIEGYVYVSGSLPKQKVASDGAOMVSGECLYNTMKCEERPKPR
	FSTVLEEMISEGNCFALNSFGDEVHMFDSLKEACGRAMRKQSAVTFSDAEFDNI
	ENTKEELIEEVEVYKTTGEVQEDOMPIGNLPDVDDPSAQAQCLTESASTSDSDCP
	MEQLVACERAPATKKEFEVYMDSOVNSELEIYRREGHNALISGESSESSQSO
	ELVAAGVPLSKMOMVEDSEVRARQANDHDLSNDYVYAKMPPREPEAGDQPRIK
	AKREFLWYLKCKLVAADKSTLVEIMRDFTGQFHSACETPKNACLTSEKVCENH
	CKRYRHPKSGKSAVAEFTREDWRAKLKLOKMKAAENEMSGSDSDRTPQGG
	TGYLFCDITFLMEIEPLINRLLEISFKRYORRAPRITLVDPGCGASTYVYEKAN
	VNOYVYTGREAEDLREPRKERNATATOLKRPVDSYLLNDTOSANVLHPEEAL
	MAHAKVYFCADDLSARSVYISAOGDSOQIATPINKVESITLTKSLKLEIDNVEKRLYRS
	PLDVASVLTFRKNTGYTSVYTSANPLVNSLTKVPRDGMSTYSIKIGTGYLFGES
	EKEERQYLYTKAEKNNVNTVHESQKTDINVLCKRLKATONEIYIPGRGNSPIMVYGV
	RHRRSIVYVYTKAEKLYFDLAELMSVQEGKLMHLEBEVK"
	118..4023
	/product="polymerase"
	/db_xref="GI:6066603"
	/protein_id="CA58364.1"
	/db_xref="GI:6066603"
	/translation="MDOSILSKITSDYVNAIILHTSATRVSDDLHNMCAVYDIOID
	STENAKRRKIDVKNRLNEDQIOLALPEPERRVTVSSVHRHTSMAAMKLETDVI
	FTSPKNGVIVYICGNWATHAKNDGGEVHCCPLIDPRDAQRTLTIDNRIDGA
	KVASEKAAVAAAOIKKCDLISENAKKADVANDLNGTFCQNKREDVYDHSHTGLG
	KKEVAAHISYDIDHLYDVDSAMERKRYVTKGFELPSADIIILCKKGLSVYNGY
	IIDBSIKSYFYDDPGNGYVHNLNLSMLVYTKTPVKAAGAAVYVLETEMGDMPFT
	ITDSEARWGVILHDSSTKPLANKRDLYVPLIDIRADLTDELVRELLISREFVNA
	LEVAQKNDQNTVAEGLISTYFASTNNAVYIGSARKTSKRDPKLLPMTITTLAVOE
	LOKKOKRVILGLKSKYKEELTLGILSEVHVHREPOSLVORGAGVAKMQRYSQ
	DLVGLHDVPLYLEINDRIKLGSAIKVNFSPLSFSLDEKSVLVEYERKQISDEL
	VSEKIGLIEGYVYVSGSLPKQKVASDGAOMVSGECLYNTMKCEERPKPR
	FSTVLEEMISEGNCFALNSFGDEVHMFDSLKEACGRAMRKQSAVTFSDAEFDNI
	ENTKEELIEEVEVYKTTGEVQEDOMPIGNLPDVDDPSAQAQCLTESASTSDSDCP
	MEQLVACERAPATKKEFEVYMDSOVNSELEIYRREGHNALISGESSESSQSO
	ELVAAGVPLSKMOMVEDSEVRARQANDHDLSNDYVYAKMPPREPEAGDQPRIK
	AKREFLWYLKCKLVAADKSTLVEIMRDFTGQFHSACETPKNACLTSEKVCENH
	CKRYRHPKSGKSAVAEFTREDWRAKLKLOKMKAAENEMSGSDSDRTPQGG
	TGYLFCDITFLMEIEPLINRLLEISFKRYORRAPRITLVDPGCGASTYVYEKAN
	VNOYVYTGREAEDLREPRKERNATATOLKRPVDSYLLNDTOSANVLHPEEAL
	MAHAKVYFCADDLSARSVYISAOGDSOQIATPINKVESITLTKSLKLEIDNVEKRLYRS
	PLDVASVLTFRKNTGYTSVYTSANPLVNSLTKVPRDGMSTYSIKIGTGYLFGES
	EKEERQYLYTKAEKNNVNTVHESQKTDINVLCKRLKATONEIYIPGRGNSPIMVYGV
	RHRRSIVYVYTKAEKLYFDLAELMSVQEGKLMHLEBEVK"
	118..4023

SOURCE	Human papillomavirus type 21.
ORGANISM	Human papillomavirus type 21
REFERENCE	Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
AUTHORS	Papillomavirus.
TITLE	1 (bases 1 to 7779)
JOURNAL	DeJius, H.
REFERENCE	Sequenced by Hajo DeJius, Deutsches Krebsforschungszentrum,
AUTHORS	Angewandte Tumorstudiologie, I.N.F. 506, W-6900 Heidelberg, Germany
TITLE	Unpublished
JOURNAL	2 (sites)
REFERENCE	Kremsdorff, D., Favre, M., Jablonska, S., Obalek, S., Rueda, L.A.,
AUTHORS	Lutzner, M.A., Blanchet-Bardon, C., Van Voort, P.C., and Oth, G.
TITLE	Molecular cloning and characterization of the genomes of nine newly
JOURNAL	recognized human papillomavirus types associated with
REFERENCE	epidermodysplasia verruciformis
AUTHORS	J. Virol. 52 (3), 1013-1018 (1984)
TITLE	85033930
JOURNAL	3 (sites)
REFERENCE	Kiyono, T., Hiraiwa, A. and Ishibashi, M.
AUTHORS	Differences in transforming activity and coded amino acid sequence
TITLE	among E6 genes of several papillomaviruses associated with
JOURNAL	epidermodysplasia verruciformis
REFERENCE	Virology 186 (2), 628-639 (1992)
AUTHORS	92124737
TITLE	4 (bases 1 to 7779)
JOURNAL	Farmer, A.D.
REFERENCE	Direct Submission
AUTHORS	Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
TITLE	Los Alamos National Laboratory, T-10, Mail Stop K710, Los Alamos,
JOURNAL	NM 87501, USA
COMMENT	HPV21 was originally isolated from skin warts of an
	epidermodysplasia verruciformis (EV) patient [2]. Hybridization
	assays and phylogenetic reconstructions based on DNA sequences
	indicate that HPV21 is most closely related to HPV14 and HPV20, and
	then to HPV19 and HPV25. This grouping agrees with assays of the
	degree of transforming activity of the E6 gene (these related HPV
	types had relatively low transforming activity as compared to HPV
	5, 8, and 47), and clustering of similarity of amino acids in the
	second zinc finger domain of E6 [3]. The E6 gene of HPVs 14, 21,
	and 25 can enhance the induction of anchorage independent growth of
	3Y1 cells by the HPV16 E7 gene, although again less effectively
	than that of HPVs 5, 8, and 47.
FEATURES	Location/Qualifiers
Source	1..7779
	/organism="Human papillomavirus type 21"
	/db_xref="taxon:31548"
	/note="cloned HPV-21 was obtained from the Papillomavirus
	Reference Center, Heidelberg and sequenced by Dr. H.
gene	DeJius"
	200..706
	/gene="E6"
CDS	200..706
	/gene="E6"
	/note="putative"
	/codon_start=1
	/product="transforming protein E6"
	/protein_id="AA079394.1"
	/db_xref="GI:1020171"
	/translation="MADSTDSADGSPKRRHLPEENTSSFLPEPLPATIRDLANIL
	ELPDDCAVPCNFCNGLFHLHLECEDEKSLILMKDKHCVACGVCACAAATYEYNE
	FYEYVGRDIEITGKSLFIDVACVCMKFLDSIEKLIDICGRKFFPHKVGSKGI
	CRICKHPO"
	706..1011
	/gene="E7"
CDS	706..1011
	/gene="E7"
	/note="putative"
	/codon_start=1
	/product="transforming protein E7"
	/protein_id="AA079395.1"
	/db_xref="GI:1020172"
	/translation="MIGKEVLIODIVLELNLQEPVQVDFCEEELPSBOOETEEEL
	PERFAKVVTPCGCKVKLRLFNATGFAIRTFNNLLFEELQLCCPCRCNCKRGKS"
gene	998..2809
	/gene="E1"
CDS	998..2809
	/gene="E1"
	/note="putative"
	/codon_start=1
	/product="replication protein E1"
	/protein_id="AA079396.1"
	/db_xref="GI:1020173"
	/translation="MADPKGTSKGLDEMDICVEAGCSQDVNDLEELPRDPDSDISE
	LDDNDLEQNGSRELFLHOESKESEEOLOKRLKRYLSKRAVAGLSPRESTITSPQK
	SKRLFAFODSGLECLTINFEEDSEVEPALDSQVPAEOLGDTHTYKELRASN
	KILMAKKEFEFGVPPNDLTROKSKTCQCAWVSVAAYHDDLESKOLQDQCDIT
	ILIRIGAMSLFLCEKCKNGYTHKLTAMLVNHEKQIISEPKLRNVAALFWK
	GAMSGATTPGTPPMTIAOOTTVGHSTEASAFDMSAVQAFDNNYLDIADIAVOYA
	KLAPEDSNVAVLANNOARVREVAASVREYKQGMKEMSEWIIHRIINEVEGEH
	WSTIAKFLRYDOOVNIMFLAALKMLHSVPKNCILILYGPNTGSAFTMSLILVLRG
	RLSEVNSKSFQFLOPMSECKLALIDVTDPCWIMDTYLRNGLDGHVSLDCKHAP
	MOTKFPALLTSTNINVHNEVNYRLHSRIKGEFNPMPKMDNTPEBELIDQSKMF
	FRIMNOLEISQDEDEGENESQSRFQCSANENHL"
	2751..4262
	/gene="E2"
CDS	2751..4262
	/gene="E2"
	/note="putative"
	/codon_start=1
	/product="regulatory protein E2"
	/protein_id="AA079397.1"
	/db_xref="GI:1020174"
	/translation="MENLDRFNVLDQOLMNIYESAANTIESQIETHMOTLRKEVLLY
	FAKQGVRLQYVYVPLAVSESRKAQIGMMLOLOSLEYKKEPMSLVNDISAEYF
	RSPEENHRKQGVSEVLYDNDKNANAYTMRYVYVDDDKHKSQSGNHTGYE
	MOGEFHYVYLFADDAISRYSRIGMEVNVNTEYAEVATSTTDPDGGADNSTST
	TPATTDSTSLSTRKOSQOOTNTRKGRYRRTTQTHQRRSRKSKSR
	SRSRISRSRSRSRSRSRSOSDOOYRRSGOVSILTTATTTTATNTSTG
	GRSSSTSSSKRRPRRGATIGSSGGRGRSSTSPSKSRGSESVRSTGRIS
	DVQKSIQSDSTRTNTRIGRLDLALDPPVILVNGEPTNLKCFENRAKILVAGLYKAF
	STAMSWAGDSTETLGRMLISFSEORXDFKTVKYPGVDRSYGSPDSL"
	3037..4017
	/gene="E4"
CDS	<3037..4017
	/gene="E4"
	/note="putative"
	/codon_start=1
	/protein_id="AA079398.1"
	/db_xref="GI:1020175"
	/translation="IPVQRHLEALKIISKGCQCLRFMTITIKTMLMPPGDMFT
	WMNTNGCIKVOAVSTTQAYILCKELDTTFYLLMVOYDAIELDIGIKTLIRKCLL
	SPAPHPHTHOEDKOTPPPPPPPPPTLPPSPRPBPNNSHNKTPKEBGTGGRVPG
	GGERPKRIKGDGRGSRGRGRGSGDPDPDGPGLPGGLNRLLTSRNTSDPDEKC
	PSSILPPPPPPPOPTTPBGQGGCHPPPPPPNGHDGEGLEAGAVGGGDHPPPP
	PAPNGHEESQSLGNVASTLTWTWESLFNQLVQIEVDLEDYTKLSIQ"
	4351..5913
	/gene="L2"
CDS	4351..5913
	/gene="L2"
	/note="putative"
	/codon_start=1
	/product="minor capsid protein L2"
	/protein_id="AA079399.1"
	/db_xref="GI:1020176"
	/translation="MARAKVKRUSADNIRYRCQKAGTQCPDVINKVESTTADKILQ
	YSGAVFEFGIGLSTGKGTGCTGYVPLDEGPAVRVGNAPTYRPLVDTTGPDSIT
	PVDTLNDPRTSSITPLDSTPDLPGBEVETIAETIHGPTRPDPDAVTTSTNCS
	AVLEVAPEPTPSRVNTRQYHNPSFQVITESTPTTGSSSLADHLIVTSGGQGTIG
	GSTPELELODPSPKSYSELEETPPKRTSTPQIONTIRRGGLTNRRILVOOVN
	ENPLFVRSRSRLVOFOFNDNPAFEEVYQIFEDIDIFNPDRDFIDITLGRPOYSE
	TPAGYAVRSRLKRGKTIIRTSQTOISGVOYFEDIDISTINTEPDIETQLIGHSQAT
	IVQPEVSTFIDIVNDENPLSEDSASDDLIDENEDSFSQIVVGGRRSSSTVP
	REFTTSGSYVODTKGYVAVPEEDKSTSDITITPDLPVYIHTFTDSGFTLHPS
	LSRKKRRKKYL"
	5929..7485
	/gene="L1"


```

repeat_region /rpl_type=inverted
2041..2047 /rpl_family="r3"
repeat_region /rpl_type=inverted
2062..2074 /rpl_family="r2"
repeat_region /rpl_type=inverted
2068..2074 /rpl_family="r3"
repeat_region /rpl_type=inverted
2112..2129 /rpl_family="r1"
gene /rpl_type=direct
complement(2137..3204)
gene="int"
complement(2137..3204)
CDS /function="Involved in site-specific phage integration"
/codon_start=1
/transl_table=1
/evidence=experimental
/product="integrase"
/protein_id="AAFI2706.1"
/db_xref="GI:6465906"
/translation="MANFRKRGKTWQRLSYKDNNGEYKKFEKGYKTKKEAAADE
AKRLNNHSEFNDISLYDFEKKAVKRPHTVETWRTYKRTLIDKYIKDKPIA
EITPTFYQAVLNKMSLILYROESLDKFPQIKSAMKIAVHEKVISSEFADPTAKSLA
ARPEVKYLADEYKLILAIAEKMEYTPACYLAVGMFAELIGTWSHVDDK
KEISIORTWYSLITNNEAETKNSSKRRKIPISKTKILKTKREIWHNKYDRIYIN
LSNNGIENKTIKVACRGVPHSLRHSFASLYIKGIDILTVSKLIGHENLVNLTQVYA
HOLKMEQENNVIRIKENKL"
complement(3242..3247)
-10_signal /evidence=experimental
complement(3264..3269)
-35_signal /evidence=not_experimental
complement(3331..3948)
CDS /note="orf2"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2707.1"
/db_xref="GI:6465907"
/translation="MTYIDELKEKQLDDGILTOEFSQAQSHLESSKEIVADPT
NSITSEGEKKEVEYFDKNCASGCAKIGVMTFRFLKDNHFCANCIKAGLITSPSN
OETIKQNTVLDTLFDHDKTKSSOPTIKRIKAPVVDKHAICPKCKSTDVVEMOQKK
GFSVGAAGVALTGTGICTAGFAGKKGKKOMHCNCGRFEYTK"
complement(3989..4177)
CDS /note="orf3"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2708.1"
/db_xref="GI:6465908"
/translation="MRKPSLAKSKAPRTTQYKRRVKKALLIPGCKGKMGAKPKPA
AYNKVYKKTFSIMDLRF"
complement(4201..4764)
CDS /note="orf4"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2709.1"
/db_xref="GI:6465909"
/translation="MKQKITIEOTIEIENHGDIVGDNSPIDATITNNKRTIYNP
QIATPRLTIELHLINCDIIRNDYDSTSPQEKRAKTEALIKLNNFEQGGTTEEL
YQLELVGCEPEKLTKIIVLSKSIKSWDEEVQVHYVYLDSTDEPSMNVYSIMAC
HIHKWMSLWSTLILNLSKFSNSORVI"
complement(4777..5121)
gene /gene="rap"
complement(4777..5121)
CDS /gene="rap"

```

```

-35_signal
-10_signal

Query Match      60.0%; Score 15; DB 11; Length 8219;
Best local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gttgaagcagca 24
Db 5925 GTTTGAAGCGCAGCA 5911
|||||
|||||

RESULT 43
AE004129 10881 bp DNA BCT 31-JUL-2000
LOCUS Vibrio cholerae chromosome I, section 37 of 251 of the complete
DEFINITION chromosome.
ACCESSION AE004129 AE003852
VERSION AE004129.1 GI:9654836
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae.
Vibrio cholerae.
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE
1 (bases 1 to 10881)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nieman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
1. 10881
Location/Qualifiers
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="biotype: El Tor"
115..1158
/gene="VC0415"
115..1158
/gene="VC0415"
/note="similar to SP:P13519 GB:M22055 PID:606191
PID:1128967 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1

```

/product="rod shape-determining protein Mrec"
 /protein_id="AAF93588.1"
 /db_xref="GI:9654837"
 /translation="MKKLRGMFNSNDLSIDLTANTLIYVKGQIVLDEPSVVAIROD
 KGRGKTYAANCHAKOMLGRPCNTSIRPMKGVIADEPVYTEKMLCHETIROVHDS
 VLKSPRYLVPCVCSSTOYERRAIRSRLGAGAEVLTIDEPMAAIGALIRYSEPG
 SMVIDIGGTEVAVISLNGVYSSVIGGDFDEATINVRNRSGLIGKATAEKI
 KHEIGSAVPDVOEIEVRGNLAEGVRPSFTLSNLELALOEPLTGIVASVVALE
 OCPPELASDISENGWLTGGALLKLDLRLMEETGIPVVIADPLTCVARGGKALE
 MDHMGDLFSEE"
 gene
 1215..2105
 /gene="VC0416"
 /note="similar to GB:M31792 SP:P16926 GB:M22055 PTD:146818
 PTD:146888; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="rod shape-determining protein Mrec"
 /protein_id="AAF93589.1"
 /db_xref="GI:9654838"
 /translation="MKPIRGSPSLQLRLEFAVMLASLMLADSRHGAFAHREPLNS
 LVAPLOYANDPRVLFDDFSDSISQOLRENNOLLKREVLTKSNLLRLPOYEENQ
 RLKRLGSSFTLDEKKVYEVMAVDTSPYRHQVYIDKGRVGVGEGPITINKEGIVGQ
 VTFVAHNSRVLLIDPNTAIFVQNTNRDRIYASGNCOTQIOLEHPTSTDIEVD
 MVTSGIGVYEGYPPVAIVSOVDKTRERFASIKADPVEHFRLLYLILPWEDRL
 OKYQADPQVLLLEENAAQO"
 gene
 2092..2580
 /gene="VC0417"
 /note="similar to GB:M31792 SP:P16927 PTD:146819
 PTD:606189 GB:000096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="rod shape-determining protein Mrec"
 /protein_id="AAF93590.1"
 /db_xref="GI:9654839"
 /translation="MLSSDIKRMVILASFLALLIOTIPWPGTLDLRPPPLLAVCC
 YWVIALPHRVNVSALVMGLIMDLIGSTLGRGMMSSIIYLVANPLVLENNALMO
 OATVIGLSMGLVLEIFCGYLIQNTVFNPLSLMSGVIAIILPMWFLILRRGHH
 VK"
 gene
 2523..3140
 /gene="VC0418"
 2523..3140
 /gene="VC0418"
 /note="similar to GB:L08793 SP:Q02169 GB:M96343 PTD:142853
 PTD:143161; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="maf protein"
 /protein_id="AAF93591.1"
 /db_xref="GI:9654840"
 /translation="MAVDVEFTAPSTPLACEMTISKVLASGSPRRRELLAQMGYOF
 EVVAVNVEEKRAAESPAYOYERLSRDKALAGALVAEAIVIGSDTIYVKDOVLEK
 PRPADAKRMILKLSQSOHOVMTCVTCRGTHSVYTTVEWFKTSLSEGELEAVYMS
 GEPCKAGSTIGLIGRFRVTRIGSHAVYGLPIYETDQLLKF"
 gene
 3166..4635
 /gene="VC0419"
 3166..4635
 /gene="VC0419"
 /note="similar to SP:P25537 PTD:48826 PTD:606187 GB:000096
 PTD:1789645; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="cytoplasmic axial filament protein"
 /protein_id="AAF93592.1"
 /db_xref="GI:9654841"
 /translation="MSAEILLNVPSERVMIMIGLIEHTIEREAKRGIVGNIKG
 KSVRLPGQAAAFVDIGDKAFLPHASDIYRTECVASEKQOQVVDISELIRGGOD
 IVQVVKDPLGKARLITDTITLPSRTLVFPGASHYGVSRILSEAEERRLAKIVSH
 YCDHGEFLIRTAAGANEKELTQDAALFKRLMSKVIERRSKYKNRSMLYGEIGLAOR

ILRDFVGTETLKILIDSROEENLKEFTSEYVELPAKLELYHGDKPIEDMYOTENEI
 OSLRKEVLESGGLIIDDQEAETITIDITGAFVGRNLEETIFPNNIEALIKRO
 LRLRLTGIIILIDFDMADEKRRVLTSLNLMKRVNTINGTOLGIVMEIKRQ
 TRSTEHVCTSTPCERGSGSVKVENVCETLILRETRVARADADKPFVYASPAVAD
 TLGGESEHALALEVFIGEVRIOAEPLINDEQDVYMM"
 gene
 4648..8523
 /gene="VC0420"
 4648..8523
 /gene="VC0420"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF93593.1"
 /db_xref="GI:9654842"
 /translation="MISTVTRGRTIMLWVLSILVLAATVALRIILLPQNRFOAI
 QHMLNONSVOVALADVOGVNTHPSISLDTLOAHMPDSDIOLNANASVEIEDLLO
 SIMORQPVADLTNGVLVDLKAIDMLALEONPFRSGRGVYKQDILLRLDDP
 TLKNSAILRTFPAGDLRODIEKLMQNGLRHRAEGVYSIAGINISLVSAFIDH
 GILRDVSGDEYVADKRVLPMLTRYLTKDQGIQGVYSLRNSDEMLPLIDVAMDQ
 KSELVWNGEBOETHLESGVELKPEKGMVNAHSRLRSDDELPIAKGATLESLEY
 PDEMRNLISQMLINELLPLAKLIPESQTLNHLITLTKRGTLEDVRYAKGATLESLEY
 SASLIDGIAQWELLPOVNAQOJOGSPKALIKASWDDVLYEGEFOAPLNIROG
 AVNLVMOQOYQMSLMSDQVYATPDOLALGAFRLDPDPDPYRHNAGVFOAGLE
 TWRYLPTLALGRELIDYLSIAQAOVNTAKILWGAIDQEPYRHNAGVFOAGLE
 AKAPFTAMPPTIDLDLIFENDAMYIDSRATIMEVSARITGRIPELAEGLHETI
 EYASAOQCNALIRNYMATPIVDSGAALTTOVGAPRASAAGLARILKOPVSDFGGEAKRG
 PALSNNADIDIPRMSLTSSGKIIFPNDRVSAAGLARILKOPVSDFGGEAKRG
 YAVIGIDGMEVKKLPIFYGGRWLSRYKGAHPQASVADIDNDGVGYTOLDGKADR
 GLESRYPPLKALKVKGQALIQASGNDQMSARQLQOAKYQAEITDITLTPVPLKAT
 NLVLGQSEKISIPVGHVHLSQANENLNDWLSLNEDEKPSKRSKSLASNTAPPKM
 PERVDAVKEITFAGLHMDVLDLNRKRLQMLIDLSQETIKGQNTIIEPDLSTALE
 RLHLFPLEAGEALLVDLDKQPLITDFPRKHOMQPLTTLIKQFWLQGYKIQG
 ANVDFPQGGTLILMKSIDPESGNTOLHVGWTWLTIDTOSRTQNMIDMGDNNSLMAR
 PGINSIGORAPFEITASTQMDGAPMSQVNTLOGVYDKLGKVISDVSAGARLLGLF
 SLDSITRKMQLDSVDYFDKGMADSISSGSELISGIFVNNIKMDAVAGEMTIKGLAD
 LNTRYDAEVEVVDITSGIPVLTAFVTPOTALYVLAITYISVVEVFOVNEVVK
 GPIDSPVKEKLSRSGEFKLEPKRLTPE"
 gene
 8532..9359
 /gene="VC0421"
 8532..9359
 /gene="VC0421"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF93594.1"
 /db_xref="GI:9654843"
 /translation="MORVGLVQNTSGPEISRLTYLKQVAKLAQOAGQWITPENAL
 LIGNRQYHOQAESLDHGPVQHALASIAEYGVMLIGSMPIRHAEGYTSLLMNAO
 GERVAVYDKLHMFDDVADGHORRESSEFTPEQOVVSTFPGALGSLICYDVPRPI
 LYADLRQGAQIILVPAFTAVTGOAHVEVLAARAIETQCVVIAVGQGHHPGRET
 WGHSMVISPGEVITANIGAEVQSKVEFDLATIDSVRRAMPITQHTFRTHOLENTH"
 gene
 9359..10804
 /gene="VC0422"
 9359..10804
 /gene="VC0422"
 /note="similar to SP:P46473 PTD:606183 GB:000096
 PTD:1732437 PTD:1789640; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="cltd protein"
 /protein_id="AAF93595.1"
 /db_xref="GI:9654844"
 /translation="MTNMOETALAPAGITBEDIADTLSSIATROIDVADYEQSSM
 HESIVLEBSITIKDGSFNDGCVRAVSGEKGFAVSDQOLEGLKQSAIANGIAQO
 GONTVQAPKRAOHPAFYGAHNPLESEKQOKTELLKQDAVIRTEKPLETTEVSLS
 GVHEOILVATDGTAGDIPRVLTSLVLAOKDRRERSAGGGRFGFYDLOSEA
 GQKIAFSPADATROALNLEAVAPAGTAPYVIGSSWPVYLIAHENGHLEDDPNK
 GSSYFAGVYGRKVTSDICTYVDDGTLDLROSLNVDDEGSGQITNVILIEGILKGYO
 DKLNRLMGVAPTGNGRRESYAHLPMRPMNTYMLPEHPEETIASVKGIITAPNG
 GGQVDTSGKTFVSAEAYLTENKQITRPIKGAATLIGSGIEAMQOVSMDNIALDRG

```

BASE COUNT      2677 a      2487 c      2982 g      2735 t
ORIGIN
VGCGKAGCGSVPGVGOPFLKDALTVGGTE"

Query Match      60.0%; Score 15; DB 1; Length 10881;
Best Local Similarity 100.0%; Pred. No. 1,4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 gatcgcgtttgaa 16
        |||
Db      6397 GATCGTGTGTTGAA 6411

RESULT  44
AE006532      11062 bp      DNA      BCT      13-APR-2001
LOCUS      Streptococcus pyogenes strain SF370 serotype M1, section 61 of 167
DEFINITION      of the complete genome.
ACCESSION      AE006532 AE004092
VERSION      AE006532.1 GI:13621984
KEYWORDS
SOURCE      Streptococcus pyogenes.
ORGANISM      Streptococcus pyogenes.
REFERENCE      1 (bases 1 to 11062)
AUTHORS      Ferrerelli,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
              Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lal,H., Lin,S.,
              Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
              Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
              Complete genome sequence of an M1 strain of Streptococcus pyogenes
              Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
              11296296
              11296296

TITLE      2 (bases 1 to 11062)
JOURNAL      Ferrerelli,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
MEDLINE      Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lal,H., Lin,S.,
PUBMED      Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
              Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
              Direct Submission
              Submitted (10-APR-2001) Department of Microbiology and Immunology,
              University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
              Oklahoma City, OK 73104, USA
              Location/Qualifiers
              1. 11062
              /organism="Streptococcus pyogenes"
              /strain="SF370"
              /serotype="M1"
              /db_xref="taxon:1314"
              85. 1248
              /gene="SPY0808"
              85. 1248
              /gene="SPY0808"
              /note="Best Blastp hit = sp|P39587|YMBD_BACSU HYPOTHETICAL
              44.4 KD PROTEIN IN EPR-GALK INTERGENIC REGION
              >g1|629026|pir|1|S39674 ymbd protein - Bacillus subtilis
              >g1|413943|emb|CA51575.1| (X73124) ipa-19d [Bacillus
              subtilis] >g1|263637|emb|CB15862.1| (Z99123) alternate
              gene name: ipa-19d
              similar to hypothetical proteins [Bacillus subtilis]"
              /codon_start=1
              /transl_table=11
              /product="conserved hypothetical protein"
              /protein_id="AAK33745.1"
              /db_xref="GI:13621985"

gene
CDS
/translation="MNLKLYDSFVEKKLAGVOLLDEKDFSNIKENQOLVIVTKSNR
EGDVGFGVITDLKDFAVFVSQNAFVSKDKEMIEAFOQVPEVKGAYECRPGDPT
ETALHGEIAQETFSILENGIAYOVINEGIMGIFLDQHDVRAIYDGLAMKSLIN
LFSTAAFSVAAMGALETTSVDLAKRSRELSLAHFEHMLNLASHHFVVMVDFEF
KYARKKLIIDVIVIDPPSFARKKQTFSSRDHKLITLALDILSPKGIITASTNA
NMVYSQRFKQIIGFSGSRPESMTLQOLPSDFITNKADERSNLYLKVFTIVRE"

gene
CDS
1245. 1931
/gene="arod"
/note="SPY0809"
1245. 1931
/gene="arod"
/BC_number="4.2.1.10"
/note="Best Blastp hit = sp|P24670|AROD_SALT1
3-DEHYDROQUINATE DEHYDRATASE (3-DEHYDROQUINASE) (TYPE I
DHQASE) >g1|96689|pir|1|S15652 3-dehydroquininate dehydratase
(BC 4.2.1.10) arod - Salmonella typhi
>g1|7546223|pdb|1OFE|A Chain A. The Structure of Type I
3-dehydroquininate dehydratase from Salmonella typhi
>g1|7546224|pdb|1OFE|B Chain B. The Structure of Type I
3-dehydroquininate dehydratase from Salmonella typhi
>g1|47642|emb|CA38418.1| (X54546) 5-dehydroquininate
hydrolyase (3-dehydroquinase) [Salmonella typhi]"
/codon_start=1
/transl_table=11
/product="putative 3-dehydroquininate dehydratase"
/protein_id="AAK33746.1"
/db_xref="GI:13621986"
/translation="MRIAIVPMRPHFDEAQAIDISKEDVNLIEMRADFLPKDEIVAV
APAIEPKAGEKIEIFLTVOEGNITLISQEVYDIKEINAIYNPYIDPEYTHRS
VGEMLDLPNLIISYHNEFEPENLEAFSEMTLAPRVYKIAVMPQSDVDLDMNY
TRGFRTLNPEQDFPATISMGKIGRLSRFGVDIGSSWTVYSLDHSVGGQVTLDMKRI
IEVLEMDISN"
2025. 3191
/gene="arof"
/note="SPY0810"
2025. 3191
/gene="arof"
/note="Best Blastp hit = db|BA05375.1| (AP001512)
chorismate synthase [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative chorismate synthase"
/protein_id="AAK33747.1"
/db_xref="GI:13621987"
/translation="MRYLTAGESSGSPSLTAIIEGIPAGLTILHPADIDHELOROGG
RGARMSIETDRVQISGVRHGKTGATITLVINKDHOKLDVMAQVAKKILAEITLKLR
VKHPRPGHADLVGQIKYHNDLRALERSRARETMMVANGAVAKKILAEITLKLR
ILIFGTTITIPSKISPRELOERALSLSIVPKQEEIKRTITDKREGDTGII
ETIVQGPALGSIYVQDKKLDKLAQAVLSINAFGVEGAEFGFGKSGQVMDET
TWTFQGYGQTNHLSGFEFGMTTQGPLVVKGVAKPIPLTKPLMSVDIDSHRYKAT
VERSDPTALPAAGVIMENVATVLAKEILETFESSTMSLQKAFSDYRAVYQGF"
3252. 3593
/gene="SPY0811"
3252. 3593
/gene="SPY0811"
/note="Best Blastp hit = pir|1A69828 hypothetical protein
yhea - Bacillus subtilis >g1|2226157|emb|CA47444.1|
(Y14080) hypothetical protein [Bacillus subtilis]
>g1|2633315|emb|CB12819.1| (Z99109) yhea [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33748.1"
/db_xref="GI:13621988"
/translation="MSQETIYDVANOLERAVRLPPEYQKYLEKKAIOADVASELDE
EVAMOKKIOGMOSGOMPTAEROTSTIOELSRKTEANDDKAVFEAQALSVYSIDIER
IVFAPLKLIVK"
3814. 5166
/gene="gor"
/note="SPY0813"
3814. 5166
/gene="gor"
/BC_number="1.6.4.2"
/note="Best Blastp hit = db|BAA76640.1| (AB019579)
glutathione reductase (GR) [Streptococcus mutans]"
/codon_start=1
/transl_table=11
/product="putative glutathione reductase (GR)"

```


[illegible]

Matches	15	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	11	tttgaagcgacgacg	25						
Db	3510	TTTGAGCGCAGCAG	3524						
RESULT 45									
AE004322/c									
LOCUS									
DEFINITION		AE004322 11328 bp DNA				BCT		31-JUL-2000	
ACCESSION		AE004322				AE003852			
VERSION		AE004322.1				GI:9657104			
KEYWORDS									
SOURCE									
ORGANISM		Vibrio cholerae.							
REFERENCE		Bacteria: Proteobacteria; gamma subdivision; Vibrionaceae: Vibrio.							
AUTHORS		1 (bases 1 to 11328)							
		Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,							
		Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,							
		Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,							
		Ernolova, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,							
		Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,							
		Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,							
		Melanos, J.J., Venter, J.C. and Fraser, C.M.							
TITLE		DNA sequence of both chromosomes of the cholera pathogen Vibrio							
JOURNAL		cholerae							
MEDLINE		Nature 406 (6795), 477-483 (2000)							
REFERENCE		2 (bases 1 to 11328)							
AUTHORS		Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,							
		Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,							
		Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,							
		Ernolova, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,							
		Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,							
		Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,							
		Melanos, J.J., Venter, J.C. and Fraser, C.M.							
TITLE		Direct Submission							
JOURNAL		Submitted (14-JUN-2000) The Institute for Genomic Research, 9712							
FEATURES		Medical Center Dr, Rockville, MD 20850, USA							
source		Location/Qualifiers							
		1. .11328							
		/organism="Vibrio cholerae"							
		/strain="N16961"							
		/serotype="O1"							
		/db_xref="taxon:666"							
		/chromosome="I"							
		/note="biotype: El Tor"							
		complement(95. .286)							
		/gene="VC2521"							
		complement(95. .286)							
		/gene="VC2521"							
		/note="Identified by Glimmer2; putative"							
		/codon_start=1							
		/transl_table=1							
		/product="hypothetical protein"							
		/protein_id="AAP95663.1"							
		/db_xref="GI:9657105"							
		/translation="MAANEDVLGIKREVEYINPPYLAIIIVCHSKRQPLKLVPRLS							
		SACRYLCWMKTFICFLDAK"							
		524. .1522							
		/gene="VC2522"							
		524. .1522							
		/gene="VC2522"							
		/note="Identified by Glimmer2; putative"							
		/codon_start=1							
		/transl_table=1							
		/product="conserved hypothetical protein"							
		/protein_id="AAP95664.1"							
		/db_xref="GI:9657106"							
		/translation="MAFRIYVYVGINMLIENVFLFIGLALIVGADKLVGSAALR							
		NIGSLPYITGMTILAMGSSAPEMMYSATALE							

```

gene
CDS
/gene="VC2523"
1572..2552
/feature="VC2523"
1572..2552
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95665.1"
/db_xref="GI:9657107"
/translation="MRPHSFDTQYVAKOVIATLHALQOLEYIINDFARACAMILA
NOTGVVVMGKSGHIGKRTAATLASTGSAFVHGEASHGDLGMEERGDYLAIS
NSGESSEIALLPVLRKLRISVISMGTGNPNSNMALADIHQITPREACPIELAPTS
STTATLVMDALAAVALMQARGTAEDEALSHPGALGRLLKTLNDIMHSDALPKVA
FOALIRDALLETISOKGGMVAIVDEODTLGTFITDGLRLDLKRIIDHSIVADVMT
ROPVYAOBNILAVESGLNLMQAKRINGLMLVENNKLVCALNMHDLKAGVM"
2552..3109
/gene="VC2524"
2552..3109
/feature="VC2524"
2552..3109
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95666.1"
/db_xref="GI:9657108"
/translation="MSTVSTPLGEVEPSLEIAKQIKLICVDGVESDPLIYMGNO
GEELEKTPRTDGYGVKALMAGIEITAITTGRSOIVENRMKALGISTLIYOGDDKVA
YDLCOKLAIPEDOTGYIGDILIMPYMEKVALVCVADGHPHLLAQKANTVTHIKGH
GAVREVCDLITQARNELDVHKGLSL"
3106..3669
/gene="VC2525"
3106..3669
/feature="VC2525"
3106..3669
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95667.1"
/db_xref="GI:9657109"
/translation="MSLSRIYVLLFTIASMSLYLLGOEDSKIQVAPNILEPMEFG
ENLENISSDEGIRNIVYITSHIDHYAKSGNTLEKAPILKAVREGTIOEMEITARRI
LSKPOVLTLYDVLAKNLLPDSGEDTLITSEMSIQLSKSPFMADKPYELNGPQEFTHG
QAMKGNFADHSALYNOVGKRYETLP"
3605..4144
/gene="VC2527"
3605..4144
/feature="VC2527"
3605..4144
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95668.1"
/db_xref="GI:9657110"
/translation="MPIIRLNFITKYVKMKRSHLSLIACLASTOYALSTDRQOPV
YIIDSQOOLMKSNRYTFLGDKLQSGISINADKLIVIRNAENGKIEELIEGQNAVAT
ESOLTDGKMLTYGAKELYYKVRDELVMINKAMLSQDDSEITGSKIRRYISIQKLLA
DSGKGRVSTYVLOQAOIE"
4147..4872
/gene="VC2528"
4147..4872
/feature="VC2528"
4147..4872
/feature="identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAF95669.1"
/db_xref="GI:9657111"
/translation="MALTKAQLAKSKYKKRKYVDVSLQVESQIYGLGPNAGTTT

```

```

gene
CDS
4928..6391
/gene="VC2529"
4928..6391
/feature="VC2529"
4928..6391
/feature="similar to PID:2317738; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-54 factor"
/protein_id="AAF95670.1"
/db_xref="GI:9657112"
/translation="MRPSIQLKIGLOLAMPOLQOARILLOLSTLDLOEIOEALSN
PLEVENSDOEAISLVMTSESPREPPELEPEPOSSDLIREKSEIALEEMOTY
DEVSANTSGIALDDAPRYOGERTOTLOPIYHMOQLDTPSPVDRTIYALIDAI
DYGYITVSLSEIEBSRDDIELDEIVAKRIQDFEYVASINLODLQLLTAY
PCDTPLWEARLLLSQYIDDLGRDYKTLTKETKLEDEIREILQILODPRPSRI
ADDAEYVLPDVSVYKEOGRWLVTTNPDSVPRLKINOYADIMRONNENSYIRTNQ
EAKMLIKLSLESNRETELKAKCIVEHODHFEYGEAMKPYLINDVAMAVEHSESTIS
RVTOKYMHPRGIFELKYFEFSHVSTONGGCSRAIRALIKLVAENPAKPLSDS
KIATLLADGIGQVARTTAKYRESIGIAPSSQRKLL"
6416..6703
/gene="VC2530"
6416..6703
/feature="VC2530"
6416..6703
/feature="similar to PID:2317739; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sigma-54 modulation protein, putative"
/protein_id="AAF95671.1"
/db_xref="GI:9657113"
/translation="MQINIOGHIDILTDSMODYVSKPKDLERFDHINHOYLARVE
KLROIAEATLHVNOAIEHAHADENMYAIDSLVOKLVRQLKHKHEKLSH"
6706..7152
/gene="VC2531"
6706..7152
/feature="VC2531"
6706..7152
/feature="similar to PID:2317740; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="pns system, nitrogen regulatory IIA component"
/protein_id="AAF95672.1"
/db_xref="GI:9657114"
/translation="MOLSEVLSLDCYKSAVOCSSKRRALETISEALHTGONATELF
ECMLSRKMGSTGICNGIAIPHARMSDSNAVAVALOCIEPEFEFSIDNRVDLFLAL
LYPEDCKEHLTKLACMAERLNDKQILQLRMAASDOELYNIMVNO"
7165..8025
/gene="VC2532"
7165..8025
/feature="VC2532"
7165..8025
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95673.1"
/db_xref="GI:9657115"
/translation="MRLIVSGSGAGKSVALARVEDLGYCVQDNLPVSLTATFATQSV
OGSQONVAVSIDIRNLPKPEPSIYDVLQDLKONNDVSMFLDASKETLLKRYSETTRI

```

Query Match

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Length 11328; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

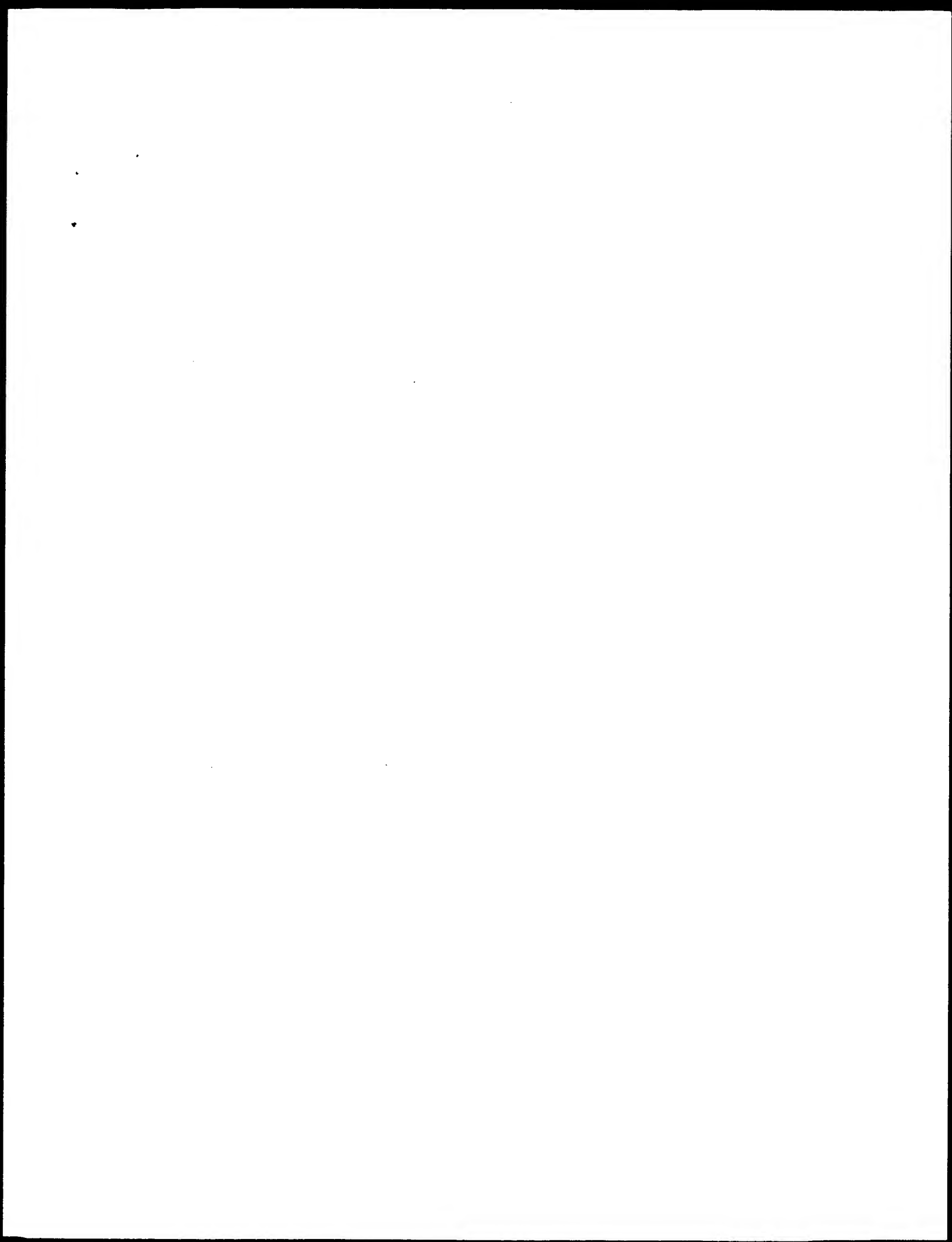
QY 1 ggaatcgtcgttga 15
 DB 7837 GGATCGCTGCTTTGA 7823

Wed Oct 10 07:45:48 2001

us-09-396-196f-4.oli.rge

Page 39

Search completed: October 9, 2001, 15:45:19
Job time: 14705 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:08 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 gcatcgtcgtttgaagcgagcagcag 25

Scoring table:

OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 9

Total number of hits satisfying chosen parameters: 36461

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: N_Geneseq_0601.*
2: /SIDSL1/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20	AAAX01303
2	25	100.0	1084	10	AAAN91329
3	25	100.0	1121	7	AAAN60496
4	25	100.0	5872	15	AAAG62386
5	17	68.0	5504	20	AAAX20559
6	16	64.0	924	22	AAAF71927
7	15	60.0	1689	20	AAAZ23024
8	15	60.0	2022	21	AAACA4132
9	15	60.0	3079	20	AAAZ23025
10	14	56.0	336	22	AAAF5487
11	14	56.0	379	20	AAAV87319

C 12	14	56.0	619	21	AAA44124
C 13	14	56.0	753	20	AAZ00347
C 14	14	56.0	1173	20	AAZ00358
C 15	14	56.0	1244	21	AAAS0896
C 16	14	56.0	1536	11	AAO04684
C 17	14	56.0	1998	20	AAZ00367
C 18	14	56.0	2087	20	AAZ00343
C 19	14	56.0	2165	20	AAZ00379
C 20	14	56.0	2405	22	AAAF27417
C 21	14	56.0	2574	20	AAZ33977
C 22	14	56.0	2574	21	AAZ38481
C 23	14	56.0	2574	21	AAZ58239
C 24	14	56.0	2669	20	AAZ25020
C 25	14	56.0	2675	21	AAZ98021
C 26	14	56.0	2778	18	AAV74359
C 27	14	56.0	12284	11	AAO06001
C 28	14	56.0	12284	15	AAO71308
C 29	14	56.0	12284	17	AAZ29591
C 30	14	56.0	63563	22	AAZ28546
C 31	13	52.0	17	19	AAO94360
C 32	13	52.0	17	19	AAV62508
C 33	13	52.0	23	16	AAO94361
C 34	13	52.0	23	19	AAV62509
C 35	13	52.0	128	21	AAAB2189
C 36	13	52.0	157	21	AAAC12134
C 37	13	52.0	192	21	AAAC11863
C 38	13	52.0	251	19	AAAC12455
C 39	13	52.0	299	21	AAAF09435
C 40	13	52.0	300	21	AAAO1371
C 41	13	52.0	384	18	AAV01888
C 42	13	52.0	391	22	AAAF65138
C 43	13	52.0	398	22	AAAF66457
C 44	13	52.0	400	21	AAAS6847
C 45	13	52.0	423	20	AAAZ9975

ALIGNMENTS

RESULT 1
ID AAX01303 standard; DNA: 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (Biot) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; Biot; ss.
XX
OS Escherichia coli.
XX
PN US5669719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
XX
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
PI WPI; 1999-152902/13.
XX
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding diaminopelargonic acid amino-transferase or biotin
PT synthase

Human secreted exp
Nucleotide sequenc
Nucleotide sequenc
Soybean PI metabol
Sali/Kpni fragment
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Human fatty acid t
Human PRO703 nucle
Human PRO703 (UNO3
Human PRO703 nucle
Human H-beta 58 fa
Human colon cancer
Staphylococcus aur
Sequence encoding
Sequence encoding
Hog cholera virus,
Genomic fragment
Septoria nodorum I
Septoria nodorum s
Septoria nodorum I
Septoria nodorum s
N. meningitidis pa
Human secreted pro
Human secreted pro
Human blallelic po
Fusarium venenatum
Human colon cancer
Human ORF27 gene.
Novel human polynu
Novel human polynu
Mycoplasma hyopneu
Fragment of human

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BIOB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 20; Length 1041;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggaatcgtctgttgaagcagcagcag 25
|||||
Db 63 ggaatcgtctgttgaagcagcagcag 87

RESULT 2

AA091329
ID AA091329 standard; DNA: 1084 BP.

AC AA091329;

DT 15-FEB-1990 (first entry)

DE E. coli Bio B gene.

KW E. coli; Bio B gene; biotin.

OS Escherichia coli.

FT Key Location/Qualifiers
CDS 24..1064
/*tag=a

GB2216530-A.

11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI: 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT - derived from E. coli and capable of replication and expression in other
PT microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
CC synthetic pathway. Pref. control sequences for expression in S. cerevisiae
CC are plasmids pMA91, pMA36c, PKV49 and pCK495, and plasmid pCK965 for
CC Lactobacillus. Insertion of bio B improves biotin yields in
CC microorganisms which export biotin, or enables growth in media contg.
CC little or no biotin of organisms unable to synthesise biotin for their
CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 10; Length 1084;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggaatcgtctgttgaagcagcagcag 25
|||||
Db 86 ggaatcgtctgttgaagcagcagcag 110

RESULT 3

AA060496
ID AA060496 standard; DNA: 1121 BP.

AC AA060496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E. coli; desthiobiotin; ds.

FT Key Location/Qualifiers
CDS 42..1082
/*tag= a

JP61149091-A.

07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI: 1986-216622/33.

DR P-PSDB; AAF60536.

XX Double stranded DNA encoding biotin synthesising enzyme -
PT comprises transformed mutant E. coli strain contg. cyclic doubled
PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
PS Disclosure: Page 534, 23pp; Japanese.

CC The sequence may be expressed by a transformed E. coli host, cultured
CC in a medium containing desthiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 7; Length 1121;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggaatcgtctgttgaagcagcagcag 25
|||||
Db 104 ggaatcgtctgttgaagcagcagcag 128

RESULT 4

AA062386
ID AA062386 standard; DNA: 5872 BP.

AC AA062386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

XX Biotin: expression; enterobacteria; vitamin H; synthesis;
KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;
KW promoter; plac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

```

FT 7-Oxononanoate aminotransferase"
FT RBS 5088..5093
FT /*tag= 0
FT /standard_name= "ORF1 RBS"
FT CDS 5098..5574
FT /*tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT /start= 5583..5644
FT terminator
FT /*tag= q
FT /standard_name= "rho-independent transcriptional
FT terminator"
FT stem_loop
FT /*tag= r
FT 5583..5605
FT
FT W09408023-A.
FT 14-APR-1994.
FT
FT PD
FT XX 01-OCT-1993; 93WO-EP02688.
FT PF
FT XX 02-OCT-1992; 92CH-0003124.
FT PR 15-JUL-1993; 93CH-0002134.
FT
FT XX (LONZ ) LONZA AG.
FT PA
FT PI Birch O., Fuhrmann M., Shaw N:
FT XX WPI: 1994-135587/16.
FT DR P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT XX Biotechnological biotin prodn. using enterobacterial biotin-gene
FT PT - providing vitamin H in high yield
FT PS
FT XX Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT XX The sequence is derived from plasmid pB030A-15/9 contg. the
FT CC bioB, bioF, bioC, bioD and bioG genes responsible for biosynthesis
FT CC of biotin, arranged in a transcription unit. Microorganisms
FT CC contg. these DNA fragments or plasmids may be used in the prodn.
FT CC of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,
FT CC loss of appetite and tiredness.
FT XX
FT SO Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY 1 ggatctgctgtttgaagcgagcagcag 25
Db 179 ggatctgctgtttgaagcgagcagcag 203
RESULT 5
AAK20559/c
ID AAK20559 standard; DNA: 5504 BP.
XX
XX AAK20559:
XX
XX 05-MAY-1999 (first entry)
XX
XX Polynucleotide sequence from the genome of Treponema pallidum.
DE
XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
XX Treponema pallidum.
OS
XX

```

PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 14-JUL-1999; 99DE-1032920.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 12-AUG-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1048613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX (BADI) BASF AG.
 XX
 XX
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberman G;
 DR WPI: 2001-13795/14.
 XX P-PSDB: AAB79808.
 XX
 PT Nucleic acids from *Corynebacterium glutamicum* encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids
 PT and purine and pyrimidine bases -
 XX
 PS
 PS Claim 3; Page 676-677; 137pp; English.
 CC AAT7153 to AAT7230 encode the *Corynebacterium glutamicum* metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carboxylic acids, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 XX Sequence 924BP: 214 A; 252 C; 239 G; 219 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 924;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cgtgttaagcgcagc 23
|||||
Db 884 CTGTTTGAAGCGCAGC 869

RESULT 7
AAZ23024/c
ID AAZ23024 standard; cDNA; 1689 BP.
XX
AC AAZ23024;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rat kd312 polypeptide encoding cDNA.
XX
KW kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; ss.
XX
OS Rattus sp.
XX
PN MO9950288-A2.
XX
PD 07-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US060993.
XX
PR 31-MAR-1998; 98US-0053374.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yen K;
XX
DR WPI; 1999-601322/51.
DR P-PSDB; AAY42694.
XX
PT kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death
XX
PS Claim 2; Fig 9; 85pp; English.
XX
CC The invention provides nucleic acid molecules encoding human and rat
CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC recombinant methodology. The kd312 sequences, and the antibodies against
CC the proteins may be used to treat or diagnose the presence or progression
CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC Parkinson's disease and Alzheimer's disease). The present sequence
CC represents the rat kd312 cDNA sequence.
XX
SQ Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1689;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatcgtcgtttga 15
|||||
Db 495 GCATCTGCTGTTGA 481

RESULT 8
AAC44132/c
ID AAC44132 standard; DNA; 2022 BP.
XX
AC AAC44132;
XX

DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

```

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142970.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154079.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154079.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 60.0%; Score 15; DB 21; Length 2022;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 tctgctgttgagc 18
Db 1268 tctgctgttgagc 1254
RESULT 9
ID AA23025/c
AA23025 standard; DNA: 3079 BP.
XX
AC AA23025;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rat Kd312 genomic DNA sequence.
XX
DE Kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; ss.

```

XX Rattus sp.
 OS WO950288-A2.
 XX 07-OCT-1999.
 PD 30-MAR-1999: 99WO-US06993.
 XX 31-MAR-1998: 98US-0053374.
 PR (AMGE-) AMGEN INC.
 XX Yuen K.
 PI WPI: 1999-601322/51.
 DR P-PSDB: AAY42694.
 XX kdl12 polypeptides useful for treating diseases and disorders
 PT associated with alterations in cell proliferation and cell death
 PS Claim 2; Fig 7; 85pp; English.
 CC The invention provides nucleic acid molecules encoding human and rat
 CC kdl12 polypeptides. The kdl12 polypeptides can be expressed by standard
 CC recombinant methodology. The kdl12 sequences, and the antibodies against
 CC the proteins may be used to treat or diagnose or progression
 CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
 CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
 CC Parkinson's disease and Alzheimer's disease). The present sequence
 CC represents the rat kdl12 genomic DNA sequence.
 SO Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gagctcgtctgttga 15
 |||||||
 DB 803 GCACTCTGCTTTGA 789

RESULT 10
 AAF65487/C
 ID AAF65487 standard; CDNA: 336 BP.
 AC AAF65487;
 XX 09-APR-2001 (first entry)
 DT
 XX Novel human polynucleotide, SEQ ID NO: 1243.
 DE
 XX Human: cytostatic; gene therapy: colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.
 XX Homo sapiens.
 OS
 PN WO200102568-A2.
 XX 11-JAN-2001.
 PD
 XX 30-JUN-2000: 2000WO-US18374.
 PF
 XX 02-JUL-1999: 99US-0142310.
 PR 02-JUL-1999: 99US-0142311.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A,
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Orkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LM, Strache-Grain B;
 XX WPI: 2001-091805/10.
 DR
 XX library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 PS Claim 9; Page 722; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;
 Query Match 56.0%; Score 14; DB 22; Length 336;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttga 16
 |||||||
 DB 159 ATCTGCTGTTGA 146

RESULT 11
 AAV87319/C
 ID AAV87319 standard; CDNA: 379 BP.
 AC AAV87319;
 XX 27-APR-1999 (first entry)
 DT
 XX EST clone BP797.
 DE
 XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX Homo sapiens.
 OS
 PN WO9845435-A2.
 XX 15-OCT-1998.
 PD
 XX 10-APR-1998: 98WO-US06954.
 PF
 XX 10-APR-1997: 97US-0835913.
 PR (GEM) GENETICS INST INC.
 XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX WPI: 1999-070076/06.
 DR New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX WPI; 1999-444398/37.
DR P-PSDB; AAY14937.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Example 1; Fig 16; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
|||||
DB 436 TCTGCTGTTGAAG 423

RESULT 14
AAZ00358/c
ID AAZ00358 standard; DNA: 1173 BP.
XX
AC AAZ00358;
XX

DT 26-OCR-1999 (first entry)
XX

DE Nucleotide sequence of human hSFATP3.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX

OS Homo sapiens.

PN W09936537-A2.
XX

PD 22-JUL-1999.
XX

PF 14-JAN-1999; 99WO-US00182.
XX

PR 14-JAN-1999; 99US-0232201.
XX

PR 15-JAN-1998; 98US-0071374.
XX

PR 20-JUL-1998; 98US-0093491.
XX

PR 04-DEC-1998; 98US-0110941.
XX

PR 14-JAN-1999; 99US-0232195.
XX

PR 14-JAN-1999; 99US-0232197.
XX

PR 14-JAN-1999; 99US-0232200.
XX

PA (MILL-) MILLENNIUM PHARM INC.
XX

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX WPI; 1999-444398/37.
DR P-PSDB; AAY14948.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Claim 31; Fig 48; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
|||||
DB 853 TCTGCTGTTGAAG 840

RESULT 15
AAA50896/c
ID AAA50896 standard; cDNA: 1244 BP.
XX

AC AAA50896;
XX

DT 28-SEP-2000 (first entry)
XX

DE Soybean PI metabolism protein clone, ssm.pk0068.c5 partial cDNA.

XX Enzyme; phosphatidylinositol-4-phosphate kinase; PI;
KW phosphatidylinositol metabolism; signal transduction; membrane anchor;
KW membrane trafficking pathway regulator; vacuole homeostasis regulator;
KW transgenic plant; soybean; EST; expressed sequence tag; ss.
XX

OS Glycine max.

PN W0200036119-A2.
XX

PD 22-JUN-2000.
XX

PF 17-DEC-1999; 99WO-US30182.
XX

PR 18-DEC-1998; 98US-0112925.
XX

PR (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX

PI Allen SM, Kinney AJ, Miao G, Rafalski JA, Sakai H, Weng Z;
 XX WPI: 2000-431597/37.
 DR P-PSDB: AAY9751.
 XX
 PT New isolated polynucleotide encoding phosphatidylinositol metabolism
 PT protein is useful for producing transgenic plants with an altered level
 PT of phosphatidylinositol -
 XX
 PS Claim 2: Page 50-51; 65pp; English.
 XX
 CC Phosphatidylinositol, PI, may be modified in cells to have roles in
 CC signal transduction, membrane associated protein anchoring and regulation
 CC of membrane trafficking pathways. The enzyme phosphatidylinositol 4-
 CC phosphate 5-kinase, converts phosphatidylinositol-4-phosphate to
 CC phosphatidylinositol-4,5-diphosphate. Phosphatidylinositol-4,5-
 CC diphosphate is thought to play a role in vacuole homeostasis regulation.
 CC The present sequence is the partial coding sequence of the soybean PI
 CC metabolism protein clone, ssm.pk0068.c5, isolated from an expressed
 CC sequence tag, EST. The present sequence may be used to create transgenic
 CC plants which express phosphatidylinositol-4-phosphate kinase, and hence
 CC be useful for modifying plant phosphatidylinositols.
 CC
 SQ Sequence 1244 BP; 392 A; 207 C; 274 G; 371 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1244;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gttgaagcagcagc 23
 |||||
 DB 886 GTTGAAGCGCAGC 873

RESULT 16

AA004684
 ID AA004684 standard; DNA; 1536 BP.

AC AA004684;

DT 05-OCT-1990 (first entry)

DE Salt/puri fragment of pUOI contg. haloacetate dehalogenase H-1.

KW Haloacetate dehalogenase; H-1; H-2; ss.

OS Moraxell sp. B.

XX Key Location/Qualifiers
 FT 413..1296
 FT /*tag= a
 FT /label=H-1_gene

PN JP02109981-A.

PD 23-APR-1990.

PF 08-APR-1988; 88JP-0085393.

PR 08-APR-1988; 88JP-0085393.

PA (MTTN) MITSUBISHI GAS CHEM KK.

WPI: 1990-168356/22.

DR P-PSDB: AAR06435.

PT Haloacetate dehalogenase coding gene - used to raise
 PT decomposition activity of aryl halogenated cpds.

PS Claim 1; Fig 2; 8pp; Japanese.

CC The DNA can be ligated into a vector which can then be used to

CC transform E.coli for the prodn. of H-1. The protein is used in
 CC the decomposition of organic halogenated cpds.
 CC See also AA004685.

SO Sequence 1536 BP; 315 A; 445 C; 449 G; 326 T; 1 other;

Query Match 56.0%; Score 14; DB 11; Length 1536;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagcagcag 25
 |||||
 DB 320 ttgaagcagcagcag 333

RESULT 17

AAZ00367/c
 ID AAZ00367 standard; DNA; 1998 BP.

AC AAZ00367;

DT 26-OCT-1999 (first entry)

DE Nucleotide sequence of murine mmFATP3.

KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

OS Mus sp.

PN W0936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99MO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

WPI: 1999-444398/37.

DR P-PSDB: AAY14957.

PT Fatty acid transport proteins and related polynucleotides, useful
 PT for treating obesity, diabetes and heart disease

XX Example 1; Fig 66; 255pp; English.

CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the

CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.
 CC
 XX
 SQ Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1998;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||
 Db 1690 TCTGCTGTTGAAG 1677

RESULT 18
 AA200343/C
 ID AA200343 standard; DNA: 2087 BP.

XX AA200343;
 XX
 AC 26-OCT-1999 (first entry)
 DE Nucleotide sequence of murine mmFATP3.
 XX
 DE Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Mus musculus.

PN W09936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Glimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

PI WPI: 1999-444398/37.

DR P-PSDB; AAY14933.

DR Fatty acid transport proteins and related polynucleotides, useful

PT for treating obesity, diabetes and heart disease

PS Example 1; Fig 8; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,

CC diabetes and heart disease.
 CC
 XX
 SQ Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2087;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||
 Db 1779 TCTGCTGTTGAAG 1766

RESULT 19
 AA200379/C
 ID AA200379 standard; DNA: 2166 BP.

XX AA200379;
 XX
 AC 26-OCT-1999 (first entry)
 DE Nucleotide sequence of human hsFATP5 gene.
 XX
 DE Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Homo sapiens.

PN W09936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Glimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

PI WPI: 1999-444398/37.

DR P-PSDB; AAY14969.

DR Fatty acid transport proteins and related polynucleotides, useful

PT for treating obesity, diabetes and heart disease

PS Claim 31; Fig 94A-B; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

```

XX Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;
SQ
Query Match
Best Local Similarity 56.0%; Score 14; DB 20; Length 2166;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgtttgaag 17
    |||
Db 1987 TCTGCTGTTGAAG 1974

RESULT 20
AAF27417/C
ID AAF27417 standard; CDNA: 2405 BP.
XX
XX AAF27417;
AC
XX
DT 24-APR-2001 (first entry)
XX
DE Human fatty acid transporter PSEC67 cDNA.
XX
KW Human; fatty acid transporter; PSEC67; long-chain fatty acid uptake;
KW oleic acid; drug screening; gene therapy; metabolic disorder;
KW cardiomyopathy; skeletal muscle disorders; renal failure; ss.
XX
OS Homo sapiens.
XX
XX MO200104301-A1.
XX
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000MO-JP04549.
XX
XX 08-JUL-1999; 99AP-0194179.
XX 18-OCT-1999; 99US-0159586.
XX 25-APR-2000; 2000JP-0128993.
XX
XX (HELI-) HELIX RES INSTR.
XX
PI Morikawa N, Masuno Y, Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
XX WPI; 2001-138349/14.
XX
XX P-PSDB; AAB60388.
XX
XX Fatty acid transporter protein and encoded gene PSEC67 cloned from
XX human cDNA library, with activity of oleic acid incorporation, useful
XX as target molecule of preventives or remedies of fatty-acid metabolic
XX disorders.
XX
XX Claim 1; Page 44-48; 58pp; Japanese.
XX
XX The invention relates to a novel human fatty acid transporter,
XX PSEC67 (AAB60388), and to cDNA encoding it (AAF27417). PSEC67 is
XX responsible for the uptake of oleic acid into cells. The invention
XX also relates to vectors and host cells comprising a PSEC67
XX nucleic acid; the recombinant production of PSEC67; an antibody
XX against PSEC67; methods of screening for compounds which can regulate
XX the uptake of long-chain fatty acids into cells; and the compounds thus
XX identified. The PSEC67 protein and the gene encoding it are useful as
XX targets for the treatment or prevention of diseases associated with
XX defective metabolism of long-chain fatty acids. Such diseases include
XX cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
XX nucleic acids may also be used in gene therapy for such disorders.
XX The present sequence represents cDNA encoding PSEC67.
XX
SQ Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;
Query Match
Best Local Similarity 56.0%; Score 14; DB 22; Length 2405;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 tctgctgtttgaag 17
    |||
Db 2111 TCTGCTGTTGAAG 2098

RESULT 21
AAZ33977/C
ID AAZ33977 standard; CDNA: 2574 BP.
XX
XX AAZ33977;
AC
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO703 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
XX Homo sapiens.
XX
XX MO9946281-A2.
XX
XX 16-SEP-1999.
XX
XX 08-MAR-1999; 99MO-US05028.
XX
XX 10-MAR-1998; 98US-0077450.
XX 11-MAR-1998; 98US-0077632.
XX 11-MAR-1998; 98US-0077641.
XX 12-MAR-1998; 98US-0077649.
XX 12-MAR-1998; 98US-0077791.
XX 13-MAR-1998; 98US-0078004.
XX 17-MAR-1998; 98US-0040220.
XX 20-MAR-1998; 98US-0078886.
XX 20-MAR-1998; 98US-0078910.
XX 20-MAR-1998; 98US-0078936.
XX 20-MAR-1998; 98US-0078939.
XX 25-MAR-1998; 98US-0079294.
XX 26-MAR-1998; 98US-0079656.
XX 27-MAR-1998; 98US-0079663.
XX 27-MAR-1998; 98US-0079664.
XX 27-MAR-1998; 98US-0079689.
XX 27-MAR-1998; 98US-0079728.
XX 27-MAR-1998; 98US-0079786.
XX 30-MAR-1998; 98US-0079920.
XX 30-MAR-1998; 98US-0079923.
XX 31-MAR-1998; 98US-0080105.
XX 31-MAR-1998; 98US-0080107.
XX 31-MAR-1998; 98US-0080165.
XX 31-MAR-1998; 98US-0080194.
XX 01-APR-1998; 98US-0080327.
XX 01-APR-1998; 98US-0080328.
XX 01-APR-1998; 98US-0080333.
XX 01-APR-1998; 98US-0080334.
XX 08-APR-1998; 98US-0081049.
XX 08-APR-1998; 98US-0081070.
XX 08-APR-1998; 98US-0081071.
XX 09-APR-1998; 98US-0081195.
XX 09-APR-1998; 98US-0081203.
XX 09-APR-1998; 98US-0081229.
XX 15-APR-1998; 98US-0081817.
XX 15-APR-1998; 98US-0081838.
XX 15-APR-1998; 98US-0081952.
XX 15-APR-1998; 98US-0081955.
XX 21-APR-1998; 98US-0082568.
XX 21-APR-1998; 98US-0082569.
XX 22-APR-1998; 98US-0082700.
XX 22-APR-1998; 98US-0082704.
XX 22-APR-1998; 98US-0082804.
XX 23-APR-1998; 98US-0082767.
XX 23-APR-1998; 98US-0082796.

```


PR 27-APR-1998: 98US-0083336.
 PR 28-APR-1998: 98US-0083332.
 PR 29-APR-1998: 98US-0083392.
 PR 29-APR-1998: 98US-0083495.
 PR 29-APR-1998: 98US-0083496.
 PR 29-APR-1998: 98US-0083499.
 PR 29-APR-1998: 98US-0083500.
 PR 29-APR-1998: 98US-0083505.
 PR 29-APR-1998: 98US-0083554.
 PR 29-APR-1998: 98US-0083558.
 PR 29-APR-1998: 98US-0083559.
 PR 30-APR-1998: 98US-0083742.
 PR 05-MAY-1998: 98US-0084366.
 PR 06-MAY-1998: 98US-0084414.
 PR 06-MAY-1998: 98US-0084441.
 PR 07-MAY-1998: 98US-0084598.
 PR 07-MAY-1998: 98US-0084600.
 PR 07-MAY-1998: 98US-0084627.
 PR 07-MAY-1998: 98US-0084637.
 PR 07-MAY-1998: 98US-0084639.
 PR 07-MAY-1998: 98US-0084640.
 PR 07-MAY-1998: 98US-0084643.
 PR 13-MAY-1998: 98US-0085323.
 PR 13-MAY-1998: 98US-0085338.
 PR 13-MAY-1998: 98US-0085339.
 PR 13-MAY-1998: 98US-0085353.
 PR 15-MAY-1998: 98US-0085579.
 PR 15-MAY-1998: 98US-0085580.
 PR 15-MAY-1998: 98US-0085582.
 PR 15-MAY-1998: 98US-0085589.
 PR 15-MAY-1998: 98US-0085697.
 PR 15-MAY-1998: 98US-0085700.
 PR 15-MAY-1998: 98US-0085704.
 PR 18-MAY-1998: 98US-0086023.
 PR 22-MAY-1998: 98US-0086032.
 PR 22-MAY-1998: 98US-0086414.
 PR 22-MAY-1998: 98US-0086430.
 PR 22-MAY-1998: 98US-0086486.
 PR 28-MAY-1998: 98US-0087098.
 PR 28-MAY-1998: 98US-0087106.
 PR 28-MAY-1998: 98US-0087208.
 PR 30-JUL-1998: 98US-0094651.
 PR 11-SEP-1998: 98US-0100038.
 XX (GETH) GENENTECH INC.
 XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 PI P-PSDB: AAY41699.
 DR WPI: 1999-551358/46.
 XX P-PSDB: AAY41699.
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 XX Claim 2: Fig 38; 530pp; English.
 PS
 XX The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AAZ33891 to
 CC AAZ3338, and AAY41685 to AAY41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;
 SQ

Query Match

56.0%; Score 14; DB 20; Length 2574;

Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 4 tctctgttgaag 17
 |||||
 Db 2167 TCTGCTGTTGAAG 2154
 RESULT 22
 AAC78481/c
 ID AAC78481 standard; cDNA; 2574 BP.
 XX
 AC AAC78481;
 XX
 XX 08-FEB-2001 (first entry)
 DE Human PRO703 (UNQ367) nucleotide sequence SEQ ID NO:101.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
 XX expressed sequence tag; detection; cancer; ss.
 OS Homo sapiens.
 XX
 PN WO200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 XX (GETH) GENENTECH INC.
 XX Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Fillaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kijavlin IU, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 DR WPI: 2000-611443/58.
 DR P-PSDB: AAB44255.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 XX Claim 2: Fig 38; 636pp; English.
 PS
 XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides can be used for detecting
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.

CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 XX

SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 2167 TCTGCTGTTGAG 2154

RESULT 23
 AAC58239/c
 ID AAC58239 standard; cDNA: 2574 BP.
 XX
 AC AAC58239;
 XX

DT 25-JAN-2001 (first entry)
 XX
 DE Human PRO703 nucleotide sequence SEQ ID NO:28.
 XX

KW Human: tumour; diagnosis; neoplastic disease; identification; cancer;
 KW tumorigenesis; detection; neoplastic cell growth; proliferation;
 KW cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
 KW immunological disorder; ss.
 XX

OS Homo sapiens.
 XX

PN W0200053754-A1.
 XX

PD 14-SEP-2000.
 XX

PF 06-JAN-2000; 2000MO-US00277.
 XX

PR 08-MAR-1999; 99MO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 05-OCT-1999; 99MO-US23089.
 PR 30-NOV-1999; 99MO-US28313.
 PR 02-DEC-1999; 99MO-US28551.
 PR 02-DEC-1999; 99MO-US28564.
 PR 30-DEC-1999; 99MO-US31243.
 PR 30-DEC-1999; 99MO-US31274.
 XX

PA (GETH) GENENTECH INC.
 XX

PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;
 PI Wood WJ;
 XX

DR WPI: 2000-572269/53.
 DR P-PSDB: AAB24054.
 XX

PT New isolated antibody for use in compositions and methods for the
 PT diagnosis and treatment of neoplastic cell growth and proliferation in
 PT mammals, including humans, and in monitoring tumor treatment -
 XX

PS Claim 50; Fig 28; 195pp; English.
 XX

CC The present invention describes an isolated antibody (Ab) that binds to
 CC one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
 CC PRO37, PRO324, PRO351, PRO362, PRO615, PRO531, PRO338, PRO3664, PRO618,
 CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
 CC and methods for the diagnosis and treatment of neoplastic cell growth
 CC and proliferation in mammals, including humans. Genes and polypeptides
 CC encoded by them, that are amplified in the genome of a tumor cell, can
 CC be identified and are useful targets for the treatment and prevention of

CC certain cancers and may be used to monitor tumour treatment. Compounds
 CC that inhibit the expression or activity of the identified polypeptides
 CC can be identified and used as antagonists. Benign or malignant tumours,
 CC inflammatory disorders and immunological disorders can be treated.
 CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
 CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
 CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 XX

SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 2167 TCTGCTGTTGAG 2154

RESULT 24
 AA25020/c
 ID AA25020 standard; cDNA: 2669 BP.
 XX
 AC AA25020;
 XX

DT 07-DEC-1999 (first entry)
 XX

DE Human H-beta 58 family CBALC05 nucleotide sequence.
 XX

KW Human: H-beta 58 family; CBALC05; cancer; AIDS; metabolic disorder;
 KW neurological disease; developmental abnormality; diagnosis;
 KW spontaneous abortion; ss.
 XX

OS Homo sapiens.
 XX

FT Key Location/Qualifiers
 FT CDS 80..1063
 FT /tag- a
 FT /product= "CBALC05"
 FT /note= "an H-beta 58 family protein"
 XX

PN W09947663-A1.
 XX

PD 23-SEP-1999.
 XX

PF 18-MAR-1998; 98MO-CN00042.
 XX

PR 18-MAR-1998; 98MO-CN00042.
 XX

PA (UYSH-) UNIV SHANGHAI SECOND MEDICAL.
 XX

PI Zhang Q, Kan L, He K, Shen Y;
 XX

DR WPI: 1999-571836/48.
 DR P-PSDB: AAY41679.
 XX

PT CBALC05, a homologue of murine H-beta 58, and related polynucleotides
 PT Claim 4; Page 9-11; 33pp; English.
 XX

CC The present sequence encodes a human H-beta 58 family protein designated
 CC CBALC05. CBALC05 is homologous to the murine H-beta 58 protein. The
 CC CBALC05 polynucleotide (1), vectors containing (1) and recombinant host
 CC cells are useful for recombinant production of CBALC05. Both (1),
 CC CBALC05 and antibodies against CBALC05 are useful as research reagents,
 CC for screening assays and in diagnostic assays. Antagonists and agonists
 CC of CBALC05 can be used to inhibit or enhance, respectively, the activity
 CC of CBALC05 or expression of (1). Anti-CBALC05 antibodies and CBALC05
 CC or its fragments can be used in vaccines. In particular, the proteins,
 CC antibodies, agonists and antagonists can be used for treating, e.g.
 CC cancer, AIDS, metabolic disorders, neurological disease, developmental

CC abnormalities and spontaneous abortions, related to both an excess and
 CC insufficient amounts of CBALC05.
 XX
 SO Sequence 2669 BP; 864 A; 419 C; 530 G; 856 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2669;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgttgaagc 18
 |||||
 Db 949 CTGCTGTTGAAGC 936

RESULT 25
 AAC98021/c
 ID AAC98021 standard; cDNA; 2675 BP.

XX AAC98021:

XX AC (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:31.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder; ss.

XX Homo sapiens.

OS WC000055351-A1.

XX 21-SEP-2000.

PD 08-MAR-2000; 2000MO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587534/55.

XX P-PSDB: AAB53264.

PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

XX Claim 1; Page 490-491; 2104pp: English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders, infectious
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

SO Sequence 2675 BP; 856 A; 425 C; 532 G; 855 T; 7 other;

Query Match 56.0%; Score 14; DB 21; Length 2675;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgttgaagc 18
 |||||
 Db 972 CTGCTGTTGAAGC 959

RESULT 26
 AAV74359/c
 ID AAV74359 standard; DNA; 7778 BP.

XX AAV74359;

XX 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #48.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

OS Key Location/Qualifiers

XX misc-feature 1681..1740

XX /tag-a

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX misc-feature 5281..5340

XX /tag-c

XX /note="these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX EP786519-A2.

XX 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.

XX 05-JAN-1996; 96US-0009861.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

XX Rosen CA;

XX WPI: 1997-374922/35.

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines

XX Claim 1: Page 387-391: 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.

SO Sequence 7778 BP; 2771 A; 1071 C; 1452 G; 2239 T; 245 other;

Query Match 56.0%; Score 14; DB 18; Length 7778;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaag 16
|||||
DB 4098 ATCTGCTGTTGAA 4085

RESULT 27
AAQ06001

ID AAQ06001 standard; DNA: 12284 BP.

XX AC AAQ06001;

DT 16-JAN-1991 (first entry)

DE Sequence encoding protein characteristic of hog cholera virus (HCV).

XX Vaccine; parvovirus; swine influenza virus.

XX Hog cholera virus.

XX Key Location/Qualifiers
FH 364..12057
FT /*tag- a

XX EP389034-A.

XX 26-SEP-1990.

XX 12-MAR-1990; 90BP-0200573.

XX 19-MAR-1989; 89BP-0104921.

XX 12-MAR-1990; 90BP-0200573.

XX (ALKU) AKZO NV.

XX Meyers G, Ruemenapf T, Thiel HJ;

XX WPI: 1990-291979/39.

XX P-PSDB; AAR06996.

XX New hog cholera virus vaccine and diagnostic - comprises nucleic
XX acid sequence of poly-peptide characteristic of hog cholera virus

PS Claim 2; Fig 2; 29pp; English.

CC Gene product may be used to provide a vaccine and Abs for diagnosis
CC of hog cholera viral infection in pigs.

SO Sequence 12284 BP; 3793 A; 2601 C; 3228 G; 2662 T; 0 other;

Query Match 56.0%; Score 14; DB 11; Length 12284;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
|||||
DB 2253 tctgctgttgaag 2266

RESULT 28
AAQ71308
ID AAQ71308 standard; DNA: 12284 BP.

XX AC AAQ71308;

DT 25-APR-1995 (first entry)

DE Sequence encoding 55 kilodalton protein of hog cholera virus.

XX Probe: detection; identification; diagnosis; hog cholera virus;

XX bovine viral diarrhoea virus; HCV; BVDV; ds.

XX Synthetic.

XX Key Location/Qualifiers
FH 364..12060
FT /*tag= a
FT /product= 55 kDa protein of Hog Cholera Virus.

XX EP614979-A.

XX 14-SEP-1994.

XX 12-MAR-1990; 90BP-0200573.

XX 19-MAR-1989; 89BP-0104921.

XX (ALKU) AKZO NOBEL NV.

XX Meyers G, Ruemenapf T, Thiel H;

XX WPI: 1994-281207/35.

XX P-PSDB; AAR60543.

XX Hog cholera virus nucleic acid and polypeptide(s) - used to
XX develop prods. for use as vaccines and for diagnosis of hog
XX cholera virus infection

PS Claim 1; Figure 2; 63pp; English.

XX Antigenic fragments of the Hog Cholera Virus (HCV) 55 kDa

XX polypeptide may be used to induce virus neutralising antibodies.

XX The antigenic fragments; cells capable of producing them and

XX recombinant HCV may be used in vaccines to protect animals against

XX HCV infection. A nucleic acid sequence comprising a vector nucleic

XX acid molecule and encoding an antigenic fragment may be used for the

XX diagnosis of HCV infection and can discriminate HCV from bovine

XX viral diarrhoea virus. The antigenic fragments can also be used in

XX diagnostic methods to detect the presence of HCV.

SO Sequence 12284 BP; 3793 A; 2600 C; 3230 G; 2661 T; 0 other;

Query Match 56.0%; Score 14; DB 15; Length 12284;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||
 Db 2253 tctgctgttgaag 2266

RESULT 29

AAAT29591
 ID AAT29591 standard; cDNA; 12284 BP.

XX AAT29591;

DF 12-AUG-1996 (first entry)

XX Hog cholera virus.

DE Hog cholera virus; HCV; swine fever; vaccine; vector; diagnosis; ds.

XX Hog cholera virus strain Alfort.

OS Hog cholera virus strain Alfort.

XX Key

PH Variation

FT Location/Qualifiers

FT 127
 FT /tag= a
 FT /note= "replace(127, 'c')"

FT CDS
 FT 364..12060
 FT /tag= b
 FT /product= 435_KDa_protein

FT variation

FT /tag= c
 FT /note= "replace(1522, 'g'), alters codon 387 from
 ACC (Thr) to GCC (Ala)"

FT variation
 FT 10989
 FT /tag= d
 FT /note= "replace(10989, 't'), alters codon 3542
 from AGG (Arg) to AGT (Ser)"

FT primer_bind
 FT complement (2587..2619)

FT primer_bind
 FT /tag= e
 FT /note= "primer 1"
 FT complement (2842..2880)

FT primer_bind
 FT /tag= f
 FT /note= "primer 2"

XX EP713915-A1.

XX 29-MAY-1996.

XX 12-MAR-1990; 90EP-0200200.

XX 19-MAR-1989; 89EP-0104921.

XX (ALKU) AKZO NOBEL NV.

XX Meyers G, Ruemenapf T, Thiel H;

XX WPI; 1996-253042/26.

XX P-PSDB; AAR95239.

XX DNA encoding hog cholera virus polypeptide - part of hog cholera
 virus precursor poly:protein, useful in vaccines and in diagnostic
 tests.

XX Claim 2; Page 12-30; 65pp: English.

XX A cDNA sequence (AAT29591) derived from the genomic RNA of hog
 cholera virus (HCV) codes for a 435 kDa polypeptide (AAR95239)

CC which, during HCV replication in an infected cell, is processed
 into the structural and non-structural proteins of the virus. The

CC cDNA library produced from PK15 and 38A1D cells infected with HCV
 strain Alfort and screened with goat anti-HCV antiserum. The cDNA,

CC or portions of it (see also AAT29594), can be used for prodn. of
 recombinant proteins useful as vaccines or diagnostics. It may

CC also be incorporated into a vector, e.g. Vaccinia virus, for use in
 vaccines, and primers based on the sequence (see also AAT29592-93)

CC are useful for diagnosis of hog cholera.
 XX Sequence 12284 BP; 3793 A; 2599 C; 3230 G; 2662 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 17; Length 12284;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||
 Db 2253 tctgctgttgaag 2266

RESULT 30

AAF28546
 ID AAF28546 standard; DNA; 63563 BP.

XX AAF28546;

DT 04-APR-2001 (first entry)

XX Genomic fragment #33.

DE Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

OS WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL.

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic

PT compositions, and for identifying virulence factors, regulatory

PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 acids -

XX Claim 1; Page 309-324; 545pp: English.

XX The present invention relates to a Moraxella catarrhalis genomic library

CC comprising of a combination of 41 nucleic acid molecules (see

CC AAF28544-AAF28554). The library has a number of uses described in the

CC specification e.g. is useful for identifying diagnostic and therapeutic

CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large

CC aerobic, gram-negative diplococcus, normally found among the bacterial

CC flora of human upper airways. M. catarrhalis is known to cause acute,

CC localised infections such as otitis media, sinusitis and bronchopulmonary

CC infection and life-threatening, systemic diseases including endocarditis

CC and meningitis.

XX Sequence 63563 BP; 18440 A; 12630 C; 11985 G; 20508 T; 0 other;

SQ

Query Match 56.0%; Score 14; DB 22; Length 63563;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatcgtcgttga 15
 |||
 Db 23739 gatcgtcgttga 23752

```

RESULT 31
AA094360/c
ID AA094360 standard; DNA: 17 BP.
XX
XX AA094360;
AC
XX
XX 04-JUN-1996 (first entry)
XX
XX
DE Septoria nodorum ITS primer JB434.
XX
XX Plant pathogen; fungus; Septoria nodorum; Septoria tritici; Fusarium;
XX Pseudocercospora herpotrichoides; Mycosphaerella fijiensis; PCR;
XX Mycosphaerella musicola; amplification; primer; ribosomal RNA gene;
XX internal transcribed region; strain; capture; colourimetric assay;
XX isolate; development; population; ss.
XX
XX Synthetic.
XX
XX MO9529260-A2.
XX
XX 02-NOV-1995.
XX
XX 19-APR-1995; 95MO-US04712.
XX
XX 25-APR-1994; 94US-0233608.
XX
XX (CIBA ) CIBA GEIGY AG.
XX
XX Beck JJ, Ligon JM;
XX
XX WPI; 1995-383005/49.
XX
XX DNA encoding intervening transcribed sequence - used for detection
XX of plant fungal pathogens
XX
XX Claim 3; Page 14; 65pp; English.
XX
XX A novel method for the detection of plant pathogenic strains of fungi
XX e.g. Septoria nodorum, S. tritici, Pseudocercospora herpotrichoides,
XX Mycosphaerella fijiensis, M. musicola or Fusarium spp. involves the PCR
XX amplification of sequences found in the internal transcribed region
XX (ITS) of the 18S, 5.8S and 28S ribosomal RNA genes by the primers
XX AA094359-93 and AAT05357-72. These primers are derived from the ITS
XX sequences of these fungi (AAT05394-105404 and AA094398) and are strain
XX specific. The amplification products of the reactions using these
XX primers can be used with the capture primers AAT05378-93 in
XX colourimetric assays. The primers and ITS DNAs can be used for the
XX detection of specific fungal pathogen isolates and in monitoring disease
XX development in plant populations.
XX
XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;
XX
XX
XX Query Match 52.0%; Score 13; DB 16; Length 17;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 12 ttgaagcgacga 24
XX |||||||||
DB 15 TTGAAGCGCAGCA 3
XX
XX
XX RESULT 32
XX AAV62508/c
XX ID AAV62508 standard; DNA: 17 BP.
XX
XX
XX AAV62508;
XX
XX
XX 17-DEC-1998 (first entry)
XX
XX Septoria nodorum species specific primer JB434.
XX
XX Internal transcribed spacer; ITS; ribosomal RNA; Fusarium avenaceum;
XX
XX Synthetic.

```

```

XX
XX Fusarium culmorum; Fusarium graminearum; Fusarium moniliforme; Plant;
XX Septoria avenae; Microdochium nivale; Fusarium poae; fungal pathogen;
XX PCR; nucleic acid detection; PCR primer; ss.
XX
XX Synthetic.
XX
XX Septoria nodorum.
XX
XX US5814453-A.
XX
XX 29-SEP-1998.
XX
XX 02-JUL-1997; 97US-0887480.
XX
XX 02-JUL-1997; 97US-0887480.
XX
XX 19-APR-1995; 95MO-US04712.
XX
XX 15-OCT-1996; 96US-0722187.
XX
XX (NOVS ) NOVARTIS FINANCE CORP.
XX
XX Beck JJ;
XX
XX WPI; 1998-541745/46.
XX
XX DNA isolated from fungal RNA, and its internal transcribed spacer
XX sequence - used for detecting fungal pathogens in plant tissue
XX
XX Example 6; Column 16; 56pp; English.
XX
XX Sequences AAV62507 to AAV62566 represent species specific PCR primers
XX for various fungal isolates used for fungal detection in the course of
XX the invention. The primers are designed based on the internal
XX transcribed spacer (ITS) sequences of the various fungal species. The
XX invention provides a DNA molecule isolated from the ribosomal RNA gene
XX region of a fungal pathogen, where the DNA molecule consists of an ITS
XX sequence selected from ITS1 and ITS2 of Fusarium culmorum, Fusarium
XX graminearum, Fusarium moniliforme, Septoria avenae or Microdochium
XX nivale. A method for detecting F. graminearum, F. culmorum, F.
XX moniliforme, F. poae, F. avenaceum and M. nivale isolates is also
XX provided which comprises isolating DNA from a plant leaf infected with at
XX least one of the above pathogens and amplifying parts of the ITS sequence
XX of the pathogen(s) by PCR using specific primers from within these
XX sequences. The pathogen(s) are detected by visualising the amplified part
XX of the ITS sequence.
XX
XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;
XX
XX
XX Query Match 52.0%; Score 13; DB 19; Length 17;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 12 ttgaagcgacga 24
XX |||||||||
DB 15 TTGAAGCGCAGCA 3
XX
XX
XX RESULT 33
XX AA094361/c
XX ID AA094361 standard; DNA: 23 BP.
XX
XX
XX AA094361;
XX
XX
XX 04-JUN-1996 (first entry)
XX
XX Septoria nodorum ITS primer JB525.
XX
XX
XX Plant pathogen; fungus; Septoria nodorum; Septoria tritici; Fusarium;
XX Pseudocercospora herpotrichoides; Mycosphaerella fijiensis; PCR;
XX Mycosphaerella musicola; amplification; primer; ribosomal RNA gene;
XX internal transcribed region; strain; capture; colourimetric assay;
XX isolate; development; population; ss.
XX
XX Synthetic.

```

```

XX PN MO9529260-A2.
XX XX 02-NOV-1995.
XX PD 19-APR-1995; 95WO-US04712.
XX PF 25-APR-1994; 94US-0233608.
XX PR (CIBA ) CIBA GEIGY AG.
XX PA Beck JJ, Ligon JM;
XX PI WPI; 1995-383005/49.
XX DR DNA encoding intervening transcribed sequence - used for detection
XX PT of plant fungal pathogens
XX PS Claim 3; Page 14; 65pp; English.
XX CC A novel method for the detection of plant pathogenic strains of fungi
XX CC e.g. Septoria nodorum, S. tritici, Pseudocercospora herpotrichoides,
XX CC Mycosphaerella filitensis, M. musicola or Fusarium spp. involves the PCR
XX CC amplification of sequences found in the internal transcribed region
XX CC (ITS) of the 18S, 5.8S and 28S ribosomal RNA genes by the primers
XX CC AAQ94359-93 and AAT05357-72. These primers are derived from the ITS
XX CC sequences of these fungi (AAT05394-r05404 and AAQ94398) and are strain
XX CC specific. The amplification products of the reactions using these
XX CC primers can be used with the capture primers AAT05378-93 in
XX CC colourimetric assays. The primers and ITS DNAs can be used for the
XX CC detection of specific fungal pathogen isolates and in monitoring disease
XX CC development in plant populations.
XX SQ Sequence 23 BP; 4 A; 6 C; 6 G; 7 T; 0 other:

Query Match 52.0%; Score 13; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
   |||||||
Db 21 TTGAAGCGCAGCA 9

RESULT 34
AAV62509/c
ID AAV62509 standard; DNA: 23 BP.
XX AC AAV62509;
XX DT 17-DEC-1998 (first entry)
XX DE Septoria nodorum species specific primer JB525.
XX KM Internal transcribed spacer; ITS; ribosomal RNA; Fusarium avenaceum;
XX KM Fusarium culmorum; Fusarium graminearum; Fusarium moniliforme; plant;
XX KM Septoria avenae; Microdochium nivale; Fusarium poae; fungal pathogen;
XX KM PCR; nucleic acid detection; PCR primer; ss.
XX OS Synthetic.
XX OS Septoria nodorum.
XX PN US5814453-A.
XX PD 29-SEP-1998.
XX PF 02-JUL-1997; 97US-0867480.
XX PR 02-JUL-1997; 97US-0867480.
XX PR 19-APR-1995; 95WO-US04712.
XX PR 15-OCT-1996; 96US-0722187.
XX

```

```

PA (NOV ) NOVARTIS FINANCE CORP.
XX Beck JJ;
XX PI WPI; 1998-541745/46.
XX DR DNA isolated from fungal RNA, and its internal transcribed spacer
XX PT sequence - used for detecting fungal pathogens in plant tissue
XX PS Example 6; Column 16; 56pp; English.
XX CC Sequences AAV62507 to AAV62566 represent species specific PCR primers
XX CC for various fungal isolates used for fungal detection in the course of
XX CC the invention. The primers are designed based on the internal
XX CC transcribed spacer (ITS) sequences of the various fungal species. The
XX CC invention provides a DNA molecule isolated from the ribosomal RNA gene
XX CC region of a fungal pathogen, where the DNA molecule consists of an ITS
XX CC sequence selected from ITS1 and ITS2 of Fusarium culmorum, Fusarium
XX CC graminearum, Fusarium moniliforme, Septoria avenae or Microdochium
XX CC nivale. A method for detecting F. graminearum, F. culmorum, F.
XX CC moniliforme, F. poae, F. avenaceum and M. nivale isolates infected with at
XX CC least one of the above pathogens and amplifying parts of the ITS sequence
XX CC of the pathogen(s) by PCR using specific primers from within these
XX CC sequences. The pathogen(s) are detected by visualising the amplified part
XX CC of the ITS sequence.
XX SQ Sequence 23 BP; 4 A; 6 C; 6 G; 7 T; 0 other:

Query Match 52.0%; Score 13; DB 19; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
   |||||||
Db 21 TTGAAGCGCAGCA 9

RESULT 35
AAAB2189
ID AAAB2189 standard; DNA: 128 BP.
XX AC AAAB2189;
XX DT 04-DEC-2000 (first entry)
XX DE N. meningitidis partial DNA sequence gnm_736 SEQ ID NO:736.
XX KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX KM Meningococcus B; MenB; ds.
XX OS Neisseria meningitidis.
XX PN WO200022430-A2.
XX PD 20-APR-2000.
XX PF 08-OCT-1999; 99WO-US23573.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 30-APR-1999; 99US-0132068.
XX PA (CHIR ) CHIRON CORP.
XX PI Frazer CM, Hickey E, Peterson J, Terteljn H, Venter JC;
XX PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX PI Rappuoli R, Pizza M;
XX DR WPI; 2000-318079/27.
XX PT Isolated nucleotide sequences of Neisseria meningitidis which can be

```


CC Libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 192 BP; 59 A; 48 C; 42 G; 43 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttga 15
 |||||
 Db 137 atctgctgttga 149

RESULT 38

AA12455 standard; DNA: 251 BP.

AA12455:

30-MAR-1999 (first entry)

Human diallelic polymorphic DNA fragment WI-19034.

DE Polymorphism; diallelic; human; forensic; paternity testing; disease;
 KM detection; phenotypic typing; characteristic; infection; hereditary;
 KM autoimmune disease; cancer; inflammation; drug; therapy; medication;
 KM treatment; marker; ss.

OS Homo sapiens.

PN W09820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; 97MO-US20313.

PR 06-NOV-1996; 96US-0030455.

PA (WHD) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI: 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics; paternity
 PT testing or phenotypic typing for disease

PS Claim 1: Page 256; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain diallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyman
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed.

CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.

SO Sequence 251 BP; 72 A; 45 C; 46 G; 87 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 251;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttga 16
 |||||
 Db 100 tctgctgttga 112

RESULT 39

AAF09435/C standard; cDNA: 299 BP.

AAF09435:

13-MAR-2001 (first entry)

Fusarium venenatum EST SEQ ID NO:1958.

DE Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.

OS Fusarium venenatum.

PN W0200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000MO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

PS Claim 86: Page 433; 3161pp; English.

XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 SQ Sequence 299 BP; 90 A; 79 C; 74 G; 56 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tttagagcgagc 23
 |||||
 Db 221 TTGAGCGCAGC 209

RESULT 40
 ID AAA01371/c
 AC AAA01371 standard; cDNA; 300 BP.

AC AAA01371;
 DT 19-MAY-2000 (first entry)

DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1362.

DE Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.

OS Homo sapiens.

PN W0958675-A2.

PD 18-NOV-1999.

PF 13-MAY-1999; 99WO-US10602.

PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085537.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.

PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
 DR WPI: 2000-126369/11.

PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells -

PS Claim 1; Page 509; 1097pp; English.

CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting

CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 SQ Sequence 300 BP; 86 A; 59 C; 57 G; 98 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ctgctgtttgaag 17
 |||||
 Db 198 CTGCTGTTGAG 186

RESULT 41
 ID AAV01888/c
 AC AAV01888 standard; cDNA; 384 BP.

AC AAV01888;

DT 20-APR-1998 (first entry)

DE Human OTK27 gene.

DE Human; foetal brain cDNA library; GDP dissociation stimulating protein;
 KW brain specific nucleosome assembly protein; diagnosis; therapy;
 KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
 KW mel-related protein type 1; mel-related type 2; hereditary disease;
 KW cancer; OTK27; ss.

OS Homo sapiens.

PN W0958675-A2.

PD 24-SEP-1997.

PF 19-MAR-1997; 97EP-0104842.

PR 05-MAR-1997; 97JP-0069163.
 PR 19-MAR-1996; 96JP-0063410.

PA (SARA) OTSUKA PHARM CO LTD.
 PI Fujiwara T, Horie M, Watanabe T;
 DR WPI: 1997-459830/43.
 DR P-PSDB: AAW37503.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -
 PT useful for diagnosis or therapy of hereditary disease and cancer
 PS Example 3; Page 35; 123pp; English.

CC The present sequence encodes a OTK27 isolated from a human foetal brain
 CC cDNA library. The nucleotide or amino acid sequences are useful for
 CC in-vitro diagnosis of hereditary diseases and cancer and for preparation
 CC of pharmaceuticals.

CC Sequence 384 BP; 94 A; 111 C; 98 G; 81 T; 0 other;

Query Match 52.0%; Score 13; DB 18; Length 384;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcacgcgcgtttc 13
 |||
 Db 349 GGATCTGCTGTTT 337

RESULT 42
 AAF65138/C
 ID AAF65138 standard; cDNA: 391 BP.

AC AAF65138;
 DT 09-APR-2001 (first entry)
 DE Novel human polynucleotide, SEQ ID NO: 894.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.
 PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
 (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 672-673; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SO Sequence 391 BP; 73 A; 118 C; 132 G; 68 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgagca 24
 |||
 Db 168 TTGAAGCGAGCA 156

RESULT 43
 AAF66457/C
 ID AAF66457 standard; cDNA: 398 BP.

AC AAF66457;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2213.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.
 PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
 (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 867; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SO Sequence 398 BP; 109 A; 76 C; 79 G; 134 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 398;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgcgtttgaag 17
 |||
 Db 254 CTGCTTTGAAG 242

```

RESULT 44
AAA56847/c
ID AAA56847 standard; DNA: 400 BP.
XX
AC AAA56847;
XX
XX 20-OCT-2000 (first entry)
XX
DE Mycoplasma hyopneumoniae genomic clone PAD983.
XX
XX Mycoplasma hyopneumoniae; vaccine; antigen; antimicrobial;
XX enzootic pneumonia; ds.
XX
OS Mycoplasma hyopneumoniae.
XX
FH Key Location/Qualifiers
FT 1..400
FT /*tag= a
FT /partial
FT /transl_except= (pos:178..180,aa:Trp)
FT /transl_except= (pos:256..258,aa:Trp)
XX
XX MO200031115-AL.
XX
XX 02-JUN-2000.
XX
XX 19-NOV-1999; 99WO-AU01035.
XX
XX 20-NOV-1998; 98AU-0007273.
XX
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
XX Moore RJ, Dorian TJ;
XX
XX WPI: 2000-400031/34.
XX
XX P-PSDB; AAB05961.
XX
XX Identifying antigenic polypeptides for use in vaccines against
XX Mycoplasma infection comprises screening proteins co-expressed with a
XX marker.
XX
XX Claim 23; Fig 28; 241pp; English.
XX
XX The present sequence was isolated from a genomic DNA library made
XX from Mycoplasma hyopneumoniae. M. hyopneumoniae causes enzootic pneumonia
XX in pigs. The infection rarely causes death but often results in
XX significant depression, leading to reduced weight gain. The animals are
XX also prone to secondary infection by opportunistic pathogens. Genomic
XX DNA from M. hyopneumoniae was partially digested with restriction
XX enzymes and fragments were ligated downstream of the polyhis tag region
XX in the expression vector pCI. The ligation mixture was used to transform
XX competent E. coli cells and polyhis positive clones were selected using
XX anti-polyhis antibodies. Expression of the recombinant vector
XX generates polyhis fusion proteins which are easily identified and
XX isolated. Pigs were vaccinated with plasmid DNA from polyhis
XX positive clones in order to identify clones encoding antigenic proteins
XX which confer protection against M. hyopneumoniae infection. Such
XX proteins may subsequently be used in the manufacture of effective
XX vaccines.
XX
SQ Sequence 400 BP; 120 A; 61 C; 63 G; 156 T; 0 other;

```

```

Query Match          52.0%; Score 13; DB 21; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 5 ctgctgtttgaag 17
   |||||||
Db 154 CTGCTGTTTGAAG 142

```

RESULT 45

```

AAx2975/c
ID AAX2975 standard; cDNA: 423 BP.
XX
AC AAX2975;
XX
XX 06-JUL-1999 (first entry)
XX
DE Fragment of human placental STAT3 coding sequence.
XX
XX Placenta; isoform; human; STAT3; intracellular; transcription factor;
XX Signal Transducer and Activator of Transcription; allele; growth arrest;
XX hepatic acute-phase protein; monocytic cell; myeloma; autoimmune disease;
XX inflammation; ss.
XX
XX Homo sapiens.
XX
XX OS
XX
XX EP906953-AL.
XX
XX 07-APR-1999.
XX
XX 16-SEP-1997; 97EP-0116061.
XX
XX 16-SEP-1997; 97EP-0116061.
XX
XX (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
XX
XX Della Pietra L, Serlupi-Crescenzi O;
XX
XX WPI: 1999-207107/18.
XX
XX P-PSDB; AAY07239.
XX
XX New allelic variant of human STAT3 useful in treating autoimmune or
XX inflammatory diseases
XX
XX Disclosure; Page 8-9; 22pp; English.
XX
XX This sequence represents a fragment of the coding sequence for the
XX placental isoform of human Signal Transducer and Activator of
XX Transcription (STAT3) intracellular transcription factor (Akira et al.,
XX Cell 77, 63-71 (1994)). The invention relates to isolation of allelic
XX variants of the placental hSTAT3 sequence. hSTAT3 plays a role in the
XX upregulation of hepatic acute-phase proteins, growth arrest of monocytic
XX cells and in the survival of myeloma cells and so may be used to treat
XX or diagnose autoimmune or inflammatory diseases.
XX
SQ Sequence 423 BP; 115 A; 106 C; 109 G; 93 T; 0 other;

```

```

Query Match          52.0%; Score 13; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

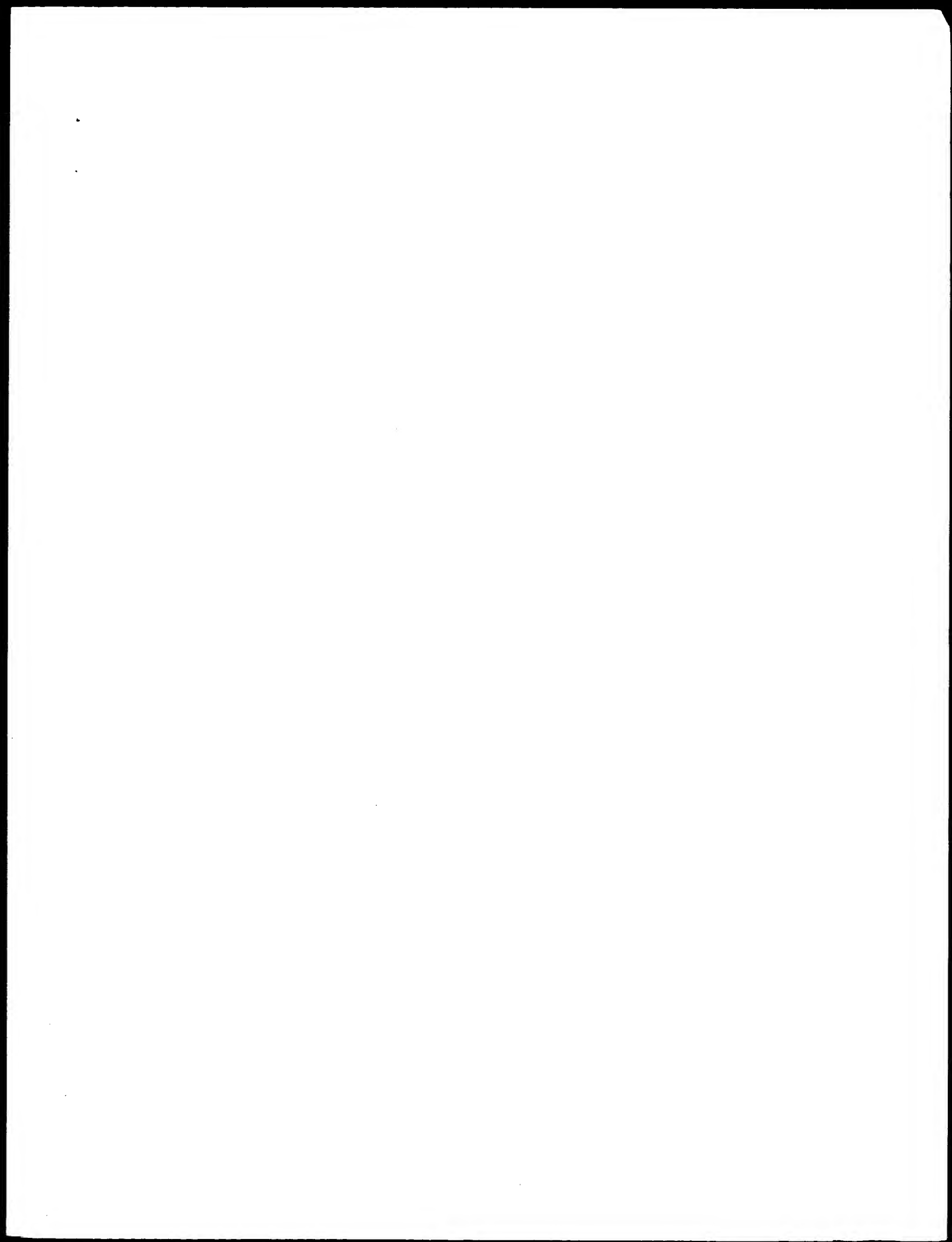
OY 13 tgaagcgcgaag 25
   |||||||
Db 142 TGAAGCGCAGCAG 130

```

Search completed: October 9, 2001, 15:52:11
Job time: 15092 sec

Wed Oct 10 07:45:49 2001

us-09-396-196f-4.oli.rng



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:26 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25
Sequence: 1 ggatctgctgttgaagcgagcagcag 25

Scoring table:
OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 11319

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	12284	2	US-08-876-991-1
6	14	56.0	12284	2	US-09-059-853-1
7	13	52.0	17	1	US-08-233-608-8
8	13	52.0	17	1	US-08-887-480-8
9	13	52.0	17	2	US-08-722-187-8
10	13	52.0	17	5	PCT-US95-04712-8
11	13	52.0	23	1	US-08-233-608-9
12	13	52.0	23	1	US-08-887-480-9
13	13	52.0	23	2	US-08-722-187-9
14	13	52.0	23	5	PCT-US95-04712-9
15	13	52.0	27	1	US-08-150-331-1
16	13	52.0	384	2	US-08-820-170A-8
17	13	52.0	384	3	US-09-055-699-8
18	13	52.0	384	4	US-09-273-565-8
19	13	52.0	583	1	US-08-233-608-2
20	13	52.0	583	1	US-08-887-480-2
21	13	52.0	583	2	US-08-722-187-2
22	13	52.0	583	5	PCT-US95-04712-2
23	13	52.0	749	4	US-08-998-416-727
24	13	52.0	1161	1	US-08-153-848-31
25	13	52.0	1161	5	PCT-US93-11153-31
26	13	52.0	1242	1	US-08-252-966B-13
27	13	52.0	1425	1	US-08-353-550-7

C 28	13	52.0	1425	2	US-08-551-687-7	Sequence 7, Appl1
C 29	13	52.0	1493	3	US-08-820-170A-9	Sequence 9, Appl1
C 30	13	52.0	1493	2	US-09-055-699-9	Sequence 9, Appl1
C 31	13	52.0	1493	4	US-09-273-565-9	Sequence 9, Appl1
C 32	13	52.0	1712	1	US-08-171-299B-1	Sequence 1, Appl1
C 33	13	52.0	1743	3	US-09-221-235-9	Sequence 9, Appl1
C 34	13	52.0	1743	3	US-09-221-928-9	Sequence 9, Appl1
C 35	13	52.0	1743	3	US-09-221-577-9	Sequence 9, Appl1
C 36	13	52.0	1743	3	US-09-221-236-9	Sequence 9, Appl1
C 37	13	52.0	1743	4	US-09-221-416-9	Sequence 9, Appl1
C 38	13	52.0	1743	4	US-09-221-245-9	Sequence 9, Appl1
C 39	13	52.0	1743	4	US-09-163-115-9	Sequence 9, Appl1
C 40	13	52.0	1743	4	US-09-221-528-9	Sequence 9, Appl1
C 41	13	52.0	1743	4	US-09-593-553-9	Sequence 9, Appl1
C 42	13	52.0	1743	4	US-09-221-237-9	Sequence 9, Appl1
C 43	13	52.0	1835	1	US-08-361-611-3	Sequence 3, Appl1
C 44	13	52.0	1835	1	US-08-565-655-3	Sequence 3, Appl1
C 45	13	52.0	1835	2	US-08-946-967-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Flimer, James Scott
REGISTRATION NUMBER: 36,129
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
|||||
DB 63 GGATCTGCTGTTTGAAGCGCAGCAG 87

RESULT 2

US-08-846-338-7

Sequence 7, Application US/08846338

Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgcagcag 25
|||||
DB 63 GGATCTGCTGTTTGAAGCGCAGCAG 87

RESULT 3

US-08-411-768B-1

Sequence 1, Application US/08411768B

Patent No. 60837712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Blass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
of Producing Biotin
NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect

SOFTWARE: Version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/411,768B

FILING DATE: 31-March-95

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: CH 3124/92

FILING DATE: 02-OCT-1992

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: CH 2134/93

FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5872 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

STRAIN: DSM498

IMMEDIATE SOURCE:

CLONE: pBO30A-15/9

FEATURE:

NAME/KEY: CDS

LOCATION: 117..1157

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 117

OTHER INFORMATION: /product= "Biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "biob"

OTHER INFORMATION: /number= 1

FEATURE:

NAME/KEY: CDS

LOCATION: 2295..3050

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 2295

OTHER INFORMATION: /function= "Involved in plimeloyl-CoA synthesis"

OTHER INFORMATION: /product= "protein"

OTHER INFORMATION: /evidence= "bioc"

OTHER INFORMATION: /gene= "bioc"

OTHER INFORMATION: /number= 3

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: aminotransf."

FEATURE:

NAME/KEY: CDS

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "bioB RBS no. 9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioC RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: Promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION: /document_number= WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gagatctgctgtttgaagcgacgacg 25
DB 179 GGATCTGCTGTTGAAGCGACGACG 203
RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTR synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "bioD RBS"
PUBLICATION INFORMATION: /document_number= WO 87/01391 B1

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttgaagcgcagcag 25
|||||
Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 5
US-08-876-991-1
Sequence 1, Application US/08876991
Patent No. 5925360
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS

LOCATION: 364..12060
OTHER INFORMATION: /label= 435_KDA_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
|||||
Db 2253 TCTGCTGTTGAAG 2266

RESULT 6
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESS: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
APPLICATION NUMBER: US 09/059,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: CDNA
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_KDA_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: replacement (127, "c")
LOCATION: replacement (1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replacement (10989, "t")
US-09-059-853-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
|||||
DB 2253 tctgctgttgaag 2266

RESULT 7
US-08-233-608-8/c
Sequence 8, Application US/08233608
Patent No. 5585238
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.608
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-233-608-8

Query Match 52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgcagca 24
|||||
DB 15 TTGAAGCGCAGCA 3

RESULT 8
US-08-887-480-8/c
Sequence 8, Application US/08887480
Patent No. 5814453
GENERAL INFORMATION:
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5814453artis Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,187
FILING DATE: 15-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8689
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: oligonucleotide primer JB434
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-887-480-8

Query Match 52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 9

US-08-722-187-8/c
; Sequence 8, Application US/08722187
; Patent No. 5955274
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,187
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US-08-722-187-8

Query Match 52.0%; Score 13; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 10

PCT-US95-04712-8/c
; Sequence 8, Application PC/TUS9504712
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04712
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-04712-8

Query Match 52.0%; Score 13; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 11

US-08-233-608-9/c
; Sequence 9, Application US/08233608
; Patent No. 5585238
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,608
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-233-608-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 1; Length 23;
Matches 13; Conservative 100.0%; Pred. No. 44;
Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 12
US-08-887-480-9/c
Sequence 9, Application US/08887480
Patent No. 5814453
GENERAL INFORMATION:
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5814453artis Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,187
FILING DATE: 15-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-887-480-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 1; Length 23;
Matches 13; Conservative 100.0%; Pred. No. 44;
Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 13
US-08-722-187-9/c
Sequence 9, Application US/08722187
Patent No. 5955274
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,187
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-722-187-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 2; Length 23;
Matches 13; Conservative 100.0%; Pred. No. 44;
Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 14
PCT-US95-04712-9/c

```

Sequence 9, Application PC/TUS9504712
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04712
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,968
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-04712-9

Query Match          52.0%; Score 13; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
DB 21 TTGAAGCGCAGCA 9

RESULT 15
US-08-150-331-31/c
Sequence 31, Application US/08150331
GENERAL INFORMATION:
APPLICANT: DORSSERS J., LAMBERTUS C.
APPLICANT: VAN LEEN, ROBERT W.
TITLE OF INVENTION: MUTANTS OF HUMAN INTERLEUKIN-3
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/150,331
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/651,437
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: GRACEY, NANCY J.
REGISTRATION NUMBER: 28,216
REFERENCE/DOCKET NUMBER: 24615-20010.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-150-331-31

```

```

Query Match          52.0%; Score 13; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
DB 20 TTGAAGCGCAGCA 8

```

```

RESULT 16
US-08-820-170A-8/c
Sequence 8, Application US/08820170A
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, MATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

MOLECULE TYPE: DNA(genomic)
US-08-820-170A-8

Query Match 52.0%; Score 13; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 17
US-09-055-699-8/c
Sequence 8, Application US/09055699
Patent No. 6005088

GENERAL INFORMATION:

APPLICANT: Tsutomu, FUJIMURA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sugitue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-09-055-699-8

Query Match 52.0%; Score 13; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 18
US-09-273-565-8/c
Sequence 8, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIMURA, TSUTOMU
APPLICANT: WATANABE, TAKESHI

APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: 0-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 8
LENGTH: 384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-273-565-8

Query Match 52.0%; Score 13; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 19
US-08-233-608-2
Sequence 2, Application US/08233608
Patent No. 5585238
GENERAL INFORMATION:
APPLICANT: Liqon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,608
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Septoria nodorum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..583
; OTHER INFORMATION: /note="DNA sequence for the
; OTHER INFORMATION: Internal transcribed Spacer of Septoria nodorum"
US-08-233-608-2

Query Match          52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
Db 464 TTGAAGCGACGA 476

RESULT 20
US-08-887-480-2
; Sequence 2, Application US/08887480
; Patent No. 5814453
; GENERAL INFORMATION:
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5814453artis Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,187
; FILING DATE: 15-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8687
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Septoria nodorum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..30
; OTHER INFORMATION: /note="3' end of small subunit
; OTHER INFORMATION: rRNA gene"
; FEATURE:
; NAME/KEY: misc_feature
; ORIGINAL SOURCE:
```

```

; LOCATION: 31..216
; OTHER INFORMATION: /note="ITS 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 217..372
; OTHER INFORMATION: /note="5.8S rRNA gene"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 373..526
; OTHER INFORMATION: /note="ITS 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527..583
; OTHER INFORMATION: /note="5' end of large subunit
; OTHER INFORMATION: rRNA gene"
US-08-887-480-2

Query Match          52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
Db 464 TTGAAGCGACGA 476

RESULT 21
US-08-722-187-2
; Sequence 2, Application US/08722187
; Patent No. 5955274
; GENERAL INFORMATION:
; APPLICANT: Iigon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,187
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```


ORGANISM: Septoria nodorum
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..583
OTHER INFORMATION: /note="DNA sequence for the
OTHER INFORMATION: Internal Transcribed Spacer of Septoria nodorum"
US-08-722-187-2

Query Match 52.0%; Score 13; DB 2; Length 583;
Best Local Similarity 100.0%; Pred No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagccagca 24
|||||
Db 464 TTGAAGCCAGCA 476

RESULT 22

PCT-US95-04712-2
Sequence 2, Application PC/TUS9504712

GENERAL INFORMATION:

APPLICANT: Ligon, James M

APPLICANT: Beck, James J

TITLE OF INVENTION: Detection of Fungal Pathogens Using the

TITLE OF INVENTION: Polymerase Chain Reaction

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04712

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/233,608

FILING DATE: 04-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: Walsh, Andrea C.

REFERENCE/DOCKET NUMBER: 34,988

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8666

TELEFAX: 919-541-8666

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 583 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Septoria nodorum

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..583

OTHER INFORMATION: /note="DNA sequence for the

OTHER INFORMATION: Internal Transcribed Spacer of Septoria nodorum"

PCT-US95-04712-2

Query Match 52.0%; Score 13; DB 5; Length 583;

Best Local Similarity 100.0%; Pred No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagccagca 24
|||||
Db 464 TTGAAGCCAGCA 476

RESULT 23

US-08-998-416-727

Sequence 727, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtel, Philipp

APPLICANT: Reibischung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwellis Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REFERENCE/DOCKET NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8687

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 727:

SEQUENCE CHARACTERISTICS:

LENGTH: 749 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: PAG1475RP

US-08-998-416-727

Query Match 52.0%; Score 13; DB 4; Length 749;
Best Local Similarity 100.0%; Pred No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcgtttgaagc 18
|||||
Db 96 TGCTGTTGAAGC 108

RESULT 24

US-08-153-848-31

Sequence 31, Application US/08153848

Query Match 52.0%; Score 13; DB 5; Length 583;

Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
US-08-153-848-31

Query Match 52.0%; Score 13; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttg 14
|||||

Db 322 GATCTGCTGTTTG 334

RESULT 25
PCT-US93-11153-31
Sequence 31, Application PC/TUS9311153
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
FEATURE:
NAME/KEY: CDS
LOCATION: 94..1158
PCT-US93-11153-31

Query Match 52.0%; Score 13; DB 5; Length 1161;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttg 14
|||||

Db 322 GATCTGCTGTTTG 334

RESULT 26
US-08-252-9668-13/C
Sequence 13, Application US/082529668
Patent No. 5624818
GENERAL INFORMATION:
APPLICANT: Eisenman, Robert N.
APPLICANT: Hurlin, Peter J.
APPLICANT: Ayer, Donald E.
TITLE OF INVENTION: Regulatory Proteins that Dimerize with
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPRLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/252,966B
 FILING DATE: 01-JUN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Shelton, Dennis K.
 REGISTRATION NUMBER: 26,997.
 REFERENCE/DOCKET NUMBER: FHC17694
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-8100
 TELEFAX: (206) 224-0779
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1242 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 DESCRIPTION: clone 10; see figure 24
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Mus musculus
 US-08-252-966B-13

Query Match 52.0%; Score 13; DB 1; Length 1242;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 cgcctgttgaag 17
 |||||||||
 DB 525 cgcctgttgaag 513

RESULT 27
 US-08-353-550-7/c
 Sequence 7, Application US/08353550
 Patent No. 5744313
 GENERAL INFORMATION:
 APPLICANT: Williams, Lewis T.
 APPLICANT: Cavanaugh, William M.
 TITLE OF INVENTION: No. 5744313el Protein Domain Which Binds
 TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend Hourie and Crew
 STREET: One Market Plaza, Stewart Tower, Suite 2000
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,550
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Dow, Karen B.
 REGISTRATION NUMBER: 29,684
 REFERENCE/DOCKET NUMBER: 02307K-057300
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-326-2400
 TELEFAX: 415-326-2422
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1425 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)

FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1422
 US-08-353-550-7

Query Match 52.0%; Score 13; DB 1; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcaca 21
 |||||||||
 DB 611 TGTTTGAAGCCCA 599

RESULT 28
 US-08-551-687-7/c
 Sequence 7, Application US/08551687
 Patent No. 5925547
 GENERAL INFORMATION:
 APPLICANT: Williams, Lewis T.
 APPLICANT: Cavanaugh, William M.
 TITLE OF INVENTION: No. 5925547el Protein Domain Which Binds
 TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew
 STREET: One Market Plaza, Stewart Tower, Suite 2000
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94105
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/551,687
 FILING DATE: 01-NOV-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/353,550
 FILING DATE: 09-DEC-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Matthew B.
 REGISTRATION NUMBER: P39,787
 REFERENCE/DOCKET NUMBER: 2307K-5731
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-326-2400
 TELEFAX: 415-326-2422
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1425 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1422
 US-08-551-687-7

Query Match 52.0%; Score 13; DB 2; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcaca 21
 |||||||||
 DB 611 TGTTTGAAGCCCA 599

```

RESULT 29
US-08-820-170A-9/C
; Sequence 9, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seass
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-08-820-170A-9

Query Match          52.0%; Score 13; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcgtcgttt 13
    |||
Db 443 GGATCGCTGTTT 431

RESULT 30
US-09-055-699-9/C
; Sequence 9, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seass
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington

```

```

; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-09-055-699-9

Query Match          52.0%; Score 13; DB 3; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcgtcgttt 13
    |||
Db 443 GGATCGCTGTTT 431

RESULT 31
US-09-273-565-9/C
; Sequence 9, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIMARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; FILE REFERENCE: O-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens

```

FEATURE:
NAME/KEY: CDS
LOCATION: (95)..(478)
US-09-273-565-9

Query Match 52.0%; Score 13; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaagcgagcag 13
|||||
Db 443 GCATCTGCTGTT 431

RESULT 32
US-08-171-299B-1
Sequence 1, Application US/08171299B
Patent No. 5599665
GENERAL INFORMATION:
APPLICANT: Bartierl, Joseph T.
TITLE OF INVENTION: EXOENZYME S PROTEIN PREPARATION AND CLONED
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171,299B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C. 35,433
REGISTRATION NUMBER: 650053,90871
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 271-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1712 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-171-299B-1

Query Match 52.0%; Score 13; DB 1; Length 1712;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgagcag 25
|||||
Db 388 TGAAGCGAGCAG 400

RESULT 33
US-09-221-235-9/c
Sequence 9, Application US/09221235
Patent No. 6043040
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,235
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-235-9

Query Match 52.0%; Score 13; DB 3; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 34
US-09-221-928-9/c
Sequence 9, Application US/09221928
Patent No. 6121030
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-928-9

Query Match 52.0%; Score 13; DB 3; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 35
US-09-221-527-9/c
Sequence 9, Application US/09221527
Patent No. 6146832
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-527-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 36
US-09-221-236-9/c
Sequence 9, Application US/09221236
Patent No. 6146841
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,236
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-236-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 37
US-09-221-416-9/c
Sequence 9, Application US/09221416
Patent No. 6153417
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,416
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)

US-09-221-416-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 38
US-09-221-245-9/c
Sequence 9, Application US/09221245
Patent No. 6180358
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,245
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: US 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-245-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
Db 64 TGAAGCGAGCAG 52

RESULT 39
US-09-163-115-9/c
Sequence 9, Application US/09163115A
Patent No. 6183962
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/163,115A
CURRENT FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-163-115-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||

Db 64 TGAAGCGCAGCAG 52

RESULT 40

US-09-221-528-9/c
; Sequence 9, Application US/09221528

; Patent No. 6190874

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/221,528

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO: 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-221-528-9

Query Match 52.0%; Score 13; DB 4; Length 1743;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

|||||

Db 64 TGAAGCGCAGCAG 52

RESULT 41

US-09-593-553-9/c

; Sequence 9, Application US/09593553

; Patent No. 6200770

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/593,553

; CURRENT FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: 09/163,115

; PRIOR FILING DATE: 1998-09-28

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO: 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-593-553-9

Query Match 52.0%; Score 13; DB 4; Length 1743;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

|||||

Db 64 TGAAGCGCAGCAG 52

RESULT 42

US-09-221-237-9/c

; Sequence 9, Application US/09221237

; Patent No. 6214597

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/221,237

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO: 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-221-237-9

Query Match 52.0%; Score 13; DB 4; Length 1743;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

|||||

Db 64 TGAAGCGCAGCAG 52

RESULT 43

US-08-361-611-3

; Sequence 3, Application US/08361611

; Patent No. 5519125

; GENERAL INFORMATION:

; APPLICANT: Potter, Sharon L.

; TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

; TITLE OF INVENTION: DNA Coding Therefor

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,611

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8614

; TELEFAX: (919) 541-8614

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1835 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 18..1469

; OTHER INFORMATION: /product= "Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"
US-08-361-611-3

Query Match 52.0%; Score 13; DB 1; Length 1835;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 15
|||||
DB 508 ATCTGCTGTTGA 520

RESULT 44

US-08-565-655-3
Sequence 3, Application US/08565655
Patent No. 5688939

GENERAL INFORMATION:

APPLICANT: Potter, Sharon L

APPLICANT: Ward, Eric R

TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

TITLE OF INVENTION: DNA Coding Therefor

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba Patent Department

STREET: 540 White Plains Rd., POB 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591-9005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/565,655

FILING DATE:

CLASSIFICATION: 210

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/361,611

FILING DATE: 12-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8614

TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1835 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 18..1469

OTHER INFORMATION: /product="Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"

US-08-565-655-3

Query Match 52.0%; Score 13; DB 1; Length 1835;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 15
|||||

DB 508 ATCTGCTGTTGA 520

RESULT 45

US-08-946-967-3

Sequence 3, Application US/08946967

Patent No. 5882869

GENERAL INFORMATION:

APPLICANT: Potter, Sharon L

APPLICANT: Ward, Eric R

TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

TITLE OF INVENTION: DNA Coding Therefor

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/946,967

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8614

TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1835 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 18..1469

OTHER INFORMATION: /product="Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"

US-08-946-967-3

Query Match 52.0%; Score 13; DB 2; Length 1835;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

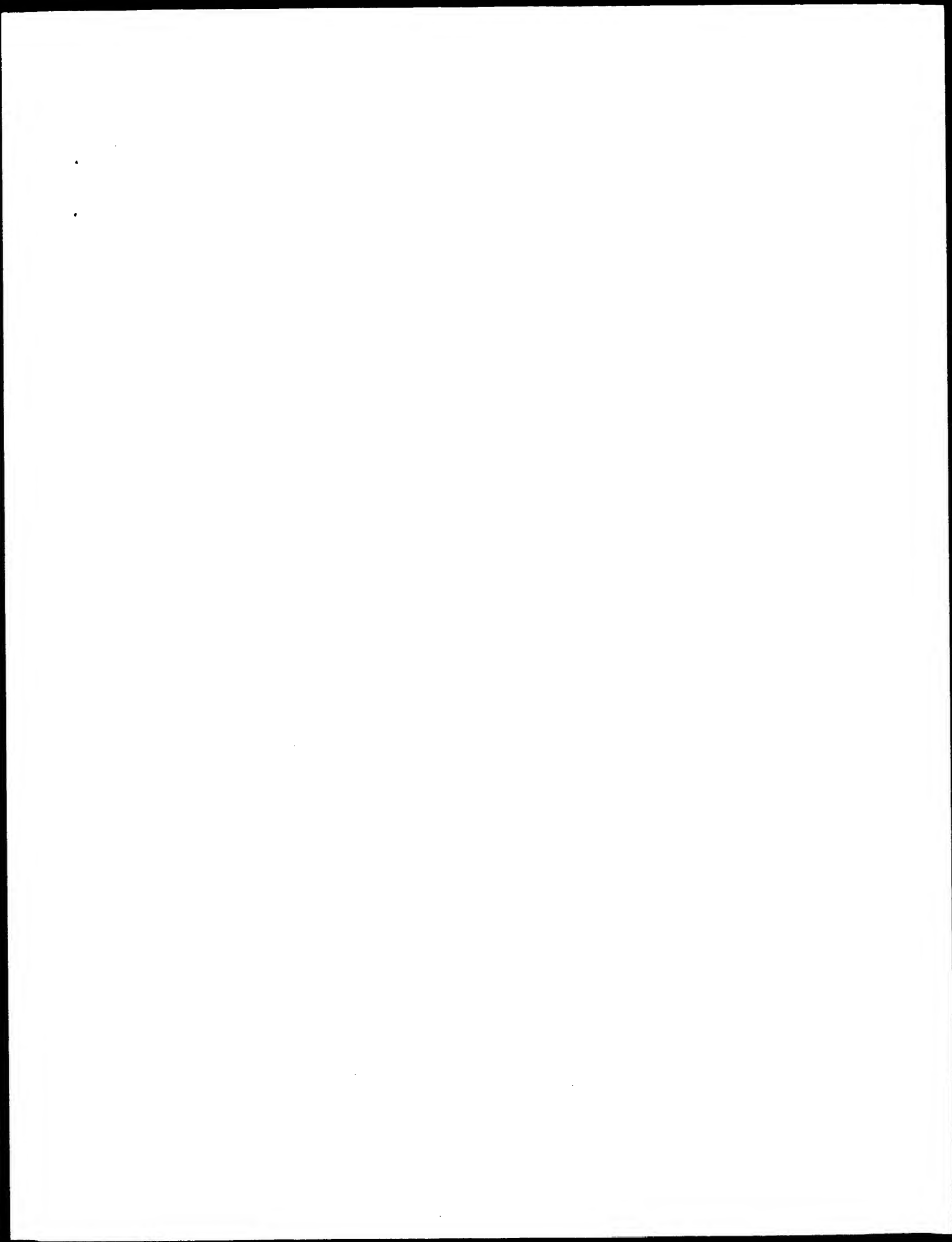
QY 3 atctgctgttga 15
|||||

DB 508 ATCTGCTGTTGA 520

Search completed: October 9, 2001, 15:55:27
Job time: 13203 sec

Wed Oct 10 07:45:50 2001

us-09-396-196f-4.oli.rni



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:26 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-4
Perfect score: 25
Sequence: 1 ggatctgctgttgaagcgacgacgacg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 656803

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	17	68.0	479	136	BE508108	BE508108 dc26f09.y
C 2	16	64.0	293	240	AZ214122	AZ214122 Sheared D
C 3	16	64.0	358	17	A1207085	A1207085 q159a09.x
C 4	16	64.0	360	156	C61070	C61070 C61070 Yuj1
C 5	16	64.0	502	136	BE515828	BE515828 WHE0606.B
C 6	16	64.0	517	232	A0692054	A0692054 HS-5408_A
C 7	16	64.0	555	235	A0889100	A0889100 HS-2204_A
C 8	16	64.0	558	111	AM064984	AM064984 ST38C02.P
C 9	16	64.0	567	165	BE232412	BE232412 137424 MA
C 10	16	64.0	632	244	AZ449344	AZ449344 1M0247J04
C 11	16	64.0	650	258	TA69C05P	AL457556 T. brucei
C 12	17	66.0	177	222	FR0044714	AL132206 Fugu rubr
C 13	15	60.0	278	15	A1021028	A1021028 ua99b01.r
C 14	15	60.0	298	131	BB337949	BB337949 BB337949
C 15	15	60.0	316	222	FR0039786	AL127281 Fugu rubr
C 16	15	60.0	358	147	BF355075	BF355075 RC2-HT082
C 17	15	60.0	372	229	A0521765	A0521765 HS-5239_A
C 18	15	60.0	406	224	A0143500	A0143500 HS-3066_B
C 19	15	60.0	424	229	A0494880	A0494880 HS-5195_A
C 20	15	60.0	435	224	A0114851	A0114851 CIT-HSP-2
C 21	15	60.0	437	116	AM432846	AM432846 sh81905.Y
C 22	15	60.0	456	150	BF555582	BF555582 UT-R-E0-C
C 23	15	60.0	466	20	A1477772	A1477772 fb54h11.Y
C 24	15	60.0	495	225	A0192174	A0192174 HS-3227_B
C 25	15	60.0	507	163	BE114310	BE114310 UT-R-E01
C 26	15	60.0	509	122	AW915326	AW915326 EST34630
C 27	15	60.0	528	122	AA818607	AA818607 UT-R-A0-b
C 28	15	60.0	532	258	TA375R11P	AL495591 T. brucei
C 29	15	60.0	542	12	AA790463	AA790463 w04g12.r
C 30	15	60.0	559	258	TA162C10P	AL472440 T. brucei
C 31	15	60.0	594	222	FR0036852	AL124355 Fugu rubr
C 32	15	60.0	598	222	FR0036852	AL123183 Fugu rubr
C 33	15	60.0	600	222	FR0044579	AL132071 Fugu rubr
C 34	15	60.0	611	222	FR0044579	AL132124 Fugu rubr
C 35	15	60.0	621	175	BC306774	BC306774 fm6b08.Y
C 36	15	60.0	627	172	BC017095	BC017095 df68g03.x
C 37	15	60.0	632	233	A0754510	A0754510 HS-5434_A
C 38	15	60.0	640	155	BC546580	BC546580 60257435
C 39	15	60.0	645	137	BE597936	BE597936 PT1-67_C0
C 40	15	60.0	669	173	BC085090	BC085090 H3108E05-
C 41	15	60.0	663	118	AM633493	AM633493 B108d10.W
C 42	15	60.0	681	140	BE820949	BE820949 GM700013A
C 43	15	60.0	748	112	AM128741	AM128741 fe3h03.Y
C 44	15	60.0	752	220	CNS02060	AL206315 Tetraodon
C 45	15	60.0	846	153	BC428569	BC428569 602501093

ALIGNMENTS

RESULT 1
BE508108/c
LOCUS
DEFINITION
sequence.
BE508108 479 bp mRNA
dc26f09.y1 NICHD XGC l11 Xenopus laevis cDNA clone IMAGE:3398249 5'
similar to TR:075940 075940 30KDA SPLICING FACTOR.; mRNA

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE508108
BE508108.1 GI:9727883
EST.
African clawed frog.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 479)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
REFERENCE
AUTHORS
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
CDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LNLN at: info@image.lnl.gov
Seq primer: -40RP from G1bco
High quality sequence stop: 434.

FEATURES

source
1. 479
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3398249"
/clone_lib="NICHD XGC l11"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMW-SPORT6; site_1: NotI;
site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."
BASE COUNT 114 a 147 c 112 g 106 t
ORIGIN

Query Match 68.0%; Score 17; DB 136; Length 479;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 tggttgaagcgagcagcagc 25
|||||
Db 183 TGTTCGAGCCGACGACG 167

RESULT 2
LOCUS
DEFINITION
A2214122 293 bp DNA GSS 09-JUN-2000
Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
A2214122
A2214122.1 GI:8431922
GSS
Trypanosoma brucei.
Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 293)
El-sayed, N., Zhao, S., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-67G1.TF
Contact: Najib M. El-sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: http://www.tigr.org/tdb/mdb/tdbd/
Seg primer: M13-Forward
Class: Shotgun.

FEATURES
source
1. 293
/organism="Trypanosoma brucei"

```

/strain="TRE0927/4 Gtat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-6761"
/clone_id="Sheared DNA"
/notes="Vector: PUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TRE0927/4 Gtat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + 1 method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds M. Vaudin and B. Barrell, Oxford University
press, 1999)."
BASE COUNT      38 a      73 c      78 g      104 t
ORIGIN

Query Match      64.0%; Score 16; DB 240; Length 293;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 atctgctgttgaagc 18
        |||
Db      149 ATCGCTGTTGAAGC 164

RESULT      3
LOCUS      AI207085      358 bp      mRNA      EST      10-NOV-1998
DEFINITION      qf59a09.x1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1754296
                 3' similar to TR:061655 Q61655 EUKARYOTIC TRANSLATION INITIATION
ACCESSION      AI207085
VERSION
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE      Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
               Email: cgaaps-r@mail.nih.gov
               CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
               , Ph.D.
               , cDNA Library Arrayed by: Greg Lennon, Ph.D.
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CCGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               www.bio.llnl.gov/bbrp/image/image.html
               Insert Length: 652 Std Error: 0.00
               Seq primer: -400P from glbco
               High quality sequence stop: 239.
               Location/Qualifiers
               1..358
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:1754296"
               /clone_id="Soares_testis_NHT"
               /sex="male"
               /lab_host="DH10B"
               /note="Vector: pT73D-Pac (Pharmacia) with a modified
               polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand CDNA
               was prepared from mRNA obtained from Clontech Laboratories
               , Inc., and primed with a Not I - Oligo(dT) primer [5'
               TGTTCACATCTGAGTGGAGCGCCCAATTTTATTTTATTTT 3'].
               Double-stranded CDNA was ligated to Eco RI adaptors
               (Pharmacia), digested with Not I and cloned into the Not I

```

```

and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      98 a      79 c      73 g      107 t      1 others
ORIGIN

Query Match      64.0%; Score 16; DB 17; Length 358;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 ctgttgaagcagc 23
        |||
Db      294 CTGTTGAAGCGCAGC 309

RESULT      4
LOCUS      C61070/c      360 bp      mRNA      EST      22-SEP-1997
DEFINITION      C61070 Yui Kohara unpublished CDNA Caenorhabditis elegans CDNA
                 clone YK224c7 5', mRNA sequence.
ACCESSION      C61070
VERSION
KEYWORDS
SOURCE      EST.
ORGANISM      Caenorhabditis elegans.
               Caenorhabditis elegans.
               Eukaryota; Metazoa; Nematoda; Chromadorea; Rhaditida; Rhaditoidae
               ; Rhaditidae; Pelodierinae; Caenorhabditis.
REFERENCE      1 (bases 1 to 360)
AUTHORS      Kohara Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
               , M., Miyata, A. and Nishigaki, A.
TITLE      Expression map of the C.elegans genome
JOURNAL      Unpublished (1996)
COMMENT      Contact: Yui Kohara
               Gene Library Lab
               National Institute of Genetics
               Yata 1111, Mishima, Shizuoka 411, Japan
               Tel: 81-559-81-6854
               Fax: 81-559-81-6855
               Email: ykohara@lab.nig.ac.jp.
               Location/Qualifiers
               1..360
               /organism="Caenorhabditis elegans"
               /strain="CB1469 him-8(e1489)"
               /db_xref="taxon:6239"
               /clone="YK224c7"
               /clone_id="Yui Kohara unpublished CDNA"
               /sex="hermaphrodite, male"
               /tissue.type="whole animal"
               /dev-stage="varied"
BASE COUNT      101 a      71 c      76 g      111 t      1 others
ORIGIN

Query Match      64.0%; Score 16; DB 156; Length 360;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 atctgctgttgaagc 18
        |||
Db      26 ATCGCTGTTGAAGC 11

RESULT      5
LOCUS      BE515828      502 bp      mRNA      EST      08-AUG-2000
DEFINITION      WHE0606_B07_C14Z4 Wheat ABA-treated embryo CDNA library Triticum
                 aestivum CDNA clone WHE0606_B07_C14, mRNA sequence.
ACCESSION      BE515828
VERSION
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum

```

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.
 1 (bases 1 to 502)
 Anderson, O.D., Chao, S., Han, P.S., Hsia, C.C., Johnson, R.R., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., Tong, J.C., Verhey, S.D., and Walker-Simmons, M.K.
 The structure and function of the expressed portion of the wheat genomes - ABA-treated embryo library
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
 Seq primer: Clontech Matchmaker 3' AD primer.
 Location/Qualifiers
 1..502
 /organism="Triticum aestivum"
 /cultivar="Brevor (soft, white, winter, common wheat)"
 /db_xref="taxon:4565"
 /clone="WHE0606_B07_C14"
 /clone_1bp="Wheat ABA-treated embryo cDNA library"
 /tissue_type="Seed embryo"
 /dev_stage="Mature dormant seeds"
 /lab_host="E. coli DH12S"
 /note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI; Embryos were cut from mature, dormant seeds and imbedded in 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7, for 12 hr at 22 C. The tissue, total RNA, and poly(A) RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA-ARS, Washington State Univ., Pullman, Washington 99164-6420. A cDNA library was made by Clontech using a combination of random and oligo dr primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
 BASE COUNT
 103 a 108 c 128 g 162 t 1 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 136; Length 502;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 atctgctgttgagc 18
 ||||||||||||||||
 Db 189 ATCTGCTGTTGAAC 204
 RESULT 6
 A0692054 517 bp DNA GSS 06-JUL-1999
 LOCUS HS_5408_A2_E07_T7A RPT-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=984 Col=14 Row=I, DNA sequence.
 ACCESSION A0692054
 VERSION A0692054.1 GI:5382302
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 517)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL
 MEDLINE
 COMMENT
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPT-11. For BAC library availability, please contact Plietier de Jong (plietier@u.washington.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm) or from Resear. h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 984 Row: I Column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 517.
 Location/Qualifiers
 1..517
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=984 Col=14 Row=I"
 /clone_1bp="RPT-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"
 BASE COUNT
 154 a 96 c 83 g 178 t 6 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 232; Length 517;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 atctgctgttgagc 18
 ||||||||||||||||
 Db 411 ATCTGCTGTTGAAC 426
 RESULT 7
 A0889100 555 bp DNA GSS 10-NOV-1999
 LOCUS HS_2204_A1_C07_MR_C1T Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
 ACCESSION A0889100
 VERSION A0889100.1 GI:6345290
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 555)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL
 MEDLINE
 COMMENT
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Resear. h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu

REFERENCE 1 (bases 1 to 632)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duyal, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A., and Wright, D., Weiss, R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0247 ROW: J Column: 04
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 632.
 Location/Qualifiers
 FEATURES
 SOURCE 1..632
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCGIM0247504"
 /clone_lib="Mouse 10kb plasmid UUCGIM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (gII47321149b/AP129072.1) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into 'chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
 BASE COUNT 185 a 167 c 110 g 170 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 244; Length 632;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ggaatcgtcgtttgaa 16
 Db 552 GCATCTGCTGTGAA 567
 RESULT 11
 LOCUS TA69C05P 650 bp DNA GSS 13-DEC-2000
 DEFINITION T. brucei sheared genomic DNA clone 69c05, forward sequence, genomic survey sequence.
 ACCESSION AL457556 GI:11858782
 VERSION AL457556.1 GI:11858782
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 650)
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
 COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREN927/4 G9rat 10.1) was mechanically sheared to give a tight size distribution (to give the library construction is 4 kb). The v + 1 method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
 Email: nh@sanger.ac.uk
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
 Location/Qualifiers
 FEATURES
 SOURCE 1..650
 /organism="Trypanosoma brucei"
 /strain="TREN927"
 /db_xref="taxon:5691"
 /clone="69c05"
 BASE COUNT 208 a 186 c 173 g 83 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 258; Length 650;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 atctcgtttgaagc 18
 Db 357 ATCTGCTGTGAAAC 342
 RESULT 12
 LOCUS FR0044714 177 bp DNA GSS 22-OCT-1999
 DEFINITION Fugu rubripes GSS sequence, clone 192G14e6, genomic survey sequence.
 ACCESSION AL132206 GI:6114152
 VERSION AL132206.1 GI:6114152
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
 1 (bases 1 to 177)
 Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umranta, Y., Williams, G. and Brenner, S.
 TITLE Direct Submission
 JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK Email: biolhel@hmp.mrc.ac.uk
 COMMENT Vector: pBluescript II KS
 V type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.
 Location/Qualifiers
 FEATURES
 SOURCE 1..177
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 192G14"


```

Query Match: 15; Score 15; DB 131; Length 298;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
OY 3 atctgctgttgaag 17
    |||||9|||||||
Db 12 ATCTGCTGTTGAAG 26

```

FEATURES	SOURCE	LOCATION/Qualifiers
BASE COUNT	85 a	1 . .316 /organization="Taki Ingu rubripes" /db_xref="taxon:31033" /clone-lib="cosmid.103002" /clone="103002B810"
ORIGIN	86 c	79 g 65 t 1 others

Query Match	60.0%;	Score 15;	DB 222;	Length 316;
Best Local Similarity	100.0%;	Pred. No. 1.1e+02;		
Matches 15; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT	16
BF355075/c	
LOCUS	BF355075 358 bp mRNA
DEFINITION	KC2-HH0987-240700-014-h01 HH0987 Homo sapiens CDNA, mRNA sequence.
ACCESSION	BF355075
VERSION	BF355075.1 GI:11314149
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE AUTHORS	TITLE
1 (Pages 1 to 356) Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zagoi, M.A., Bordi, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, P., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.J.	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL MEDLINE 20202663 Contact: Simpson A.J.G.	PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimmons@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1-RC2&t2-RC2-H10827-240700-014-h01&t3=2000-07-24&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 356

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="HT0827"
/dev_stage="Adult"
/notes="Organ: head/neck; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORESMES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

Query Match	60.0%	Score 15:	DB 147:	Length 358:
Similarity	100.0%	Pred No. 1.1e+02:		
Best Local				
Matches 15: Conservative	0:	Mismatches 0:	Indels 0:	Gaps 0:

```

Qy      4 tctgctgttgaagc 18
          |||||
Db    275 TCTGCTGTTGAAGC 261

```

RESULT	17				
LOCUS	AO521765				
DEFINITION	AO521765	372 bp	DNA	GSS	05-MAY-1999
	HS-2233-A2	H09-SP6E	RCR1-11	human Male	BAC Library Homo sapiens
	genomic clone	Plate-815	Col-18	Row-O,	DNA sequence.

ACCESSION A0521765
 VERSION A0521765.1 GI:4752890
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 372)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pliet de Jong (piet@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 815 row: 0 column: 18
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 372.
 FEATURES
 source
 Location/Qualifiers
 1..372
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=815 Col=18 Row=0"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
 BASE COUNT 97 a 86 c 72 g 115 t 2 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 229; Length 372;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 gatctgctgtttgaa 16
 ||||||||||||||||
 Db 315 GATCTGCTGTTTCAA 329
 RESULT 18
 A0143500 406 bp DNA GSS 24-SEP-1998
 LOCUS A0143500/C
 DEFINITION HS_3066_B2_H10_MR_C1T Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3066 Col=20 Row=P, DNA sequence.
 ACCESSION A0143500
 VERSION A0143500.1 GI:3534153
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 406)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pliet de Jong (piet@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 771 row: C column: 11
 Seq primer: SP6
 Class: BAC ends
 FEATURES
 source
 Location/Qualifiers
 1..406
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3066 Col=20 Row=P"
 /clone_lib="C1T Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 138 a 78 c 54 g 135 t 1 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 224; Length 406;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 gatctgctgtttgaa 16
 ||||||||||||||||
 Db 98 GATCTGCTGTTTCAA 84
 RESULT 19
 A0494880 424 bp DNA GSS 28-APR-1999
 LOCUS A0494880
 DEFINITION HS_5195_AL_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=771 Col=11 Row=C, DNA sequence.
 ACCESSION A0494880
 VERSION A0494880.1 GI:4695047
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 424)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pliet de Jong (piet@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 771 row: C column: 11
 Seq primer: SP6
 Class: BAC ends

RESULT	23
A1477772/c	
LOCUS	A1477772 466 bp mRNA
DEFINITION	fb5dhl1.y1 zebrafish WASHU MPING EST Danio rerio cDNA 5' , mRNA sequence.

```

BASE COUNT
ORIGIN
97 a 135 c 120 g 113 t 1 others
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPMG ESM"
/sex="mixed"
/lssue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dt)15 primer
[5' pGACATGTTCTAGATCGCAGCGCCGCCCTTTTCTTTTCTT3']
double-stranded cDNA was ligated to Sal I adaptors (BR),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BR). Library was constructed
by Matthew Clark (Lehrnach Lab, ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for ESM
analysis were selected following arrayed oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

```

Query Match	60.0%;	Score 15;	DB 20;	Length 466;
Best Local Similarity	100.0%;	Pred. NO. 1.1e+02;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	8	ctgtttgaagcgcag	22	
Db	330	ctgtttgaagcgcag	316	

RESULT 24
 A0192174 495 bp DNA GSS 04-NOV-1998
 LOCUS HS.3227.B1.G01.T7 CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3227 Col=1 Row=N, DNA sequence.
 ACCESSION A0192174
 VERSION A0192174.1 GI:3590796
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 495)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE Contact: Mahairas GG, Wallace JC, Hood L
 COMMENT High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3227 Row: N column: 1
 Class: BAC ends
 High quality sequence stop: 495.
 Location/Qualifiers
 1. 495
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3227 Col=1 Row=N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 135 a 114 c 79 g 167 t
 ORIGIN

Query Match 60.0%; Score 15; DB 225; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaa 16
 |||||||||||||
 DB 316 GATCTGCTTTGAA 330

RESULT 25
 BE114310 507 bp mRNA EST 13-JUN-2000
 LOCUS BE114310/c
 DEFINITION UI-R-BJ1-awk-g-02-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
 UI-R-BJ1-awk-g-02-0-UI 3, mRNA sequence.
 ACCESSION BE114310
 VERSION BE114310.1 GI:8506415
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 507)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE Contact: Soares, MB
 COMMENT

Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dt track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dt track served to identify it as a clone from the
 normalized heart library cDNA library preparation: M.B. Soares Lab
 Clone distribution: clones will be available through Research
 Genetics (www.resgen.com)
 Seq primer: M13 forward
 POLYA-Yes.

FEATURES
 source
 1. 507
 location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ1-awk-g-02-0-UI"
 /clone_lib="UI-R-BJ1"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
 library is a subtracted library derived from the following
 tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
 canal at 16.5 dpc, ventricle at 13 dpc, and adult heart.
 For a detailed description of the library from which this
 clone was derived, please visit our web site at
 ratseq.eng.uiowa.edu. The subtraction has been previously
 described in (Bonaldo, Lennon and Soares, Genome Research
 6:791-806, 1996)
 TAG_LIB=UI-R-BJ1
 TAG_TISSUE=heart
 TAG_SEQ=ACAC"
 BASE COUNT 125 a 119 c 105 g 158 t
 ORIGIN

Query Match 60.0%; Score 15; DB 163; Length 507;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgtttgaa 18
 |||||||||||||
 DB 357 TCTGCTGTTGAAC 343

RESULT 26
 AW915326 509 bp mRNA EST 25-MAY-2000
 LOCUS AW915326/c
 DEFINITION EST346630 Normalized rat embryo. Bento Soares Rattus sp. cDNA clone
 RGC011 5' end, mRNA sequence.
 ACCESSION AW915326
 VERSION AW915326.1 GI:8081017
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 509)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igf.org
 This clone is available through the ATCC, contact the ATCC
 tel#703-365-2700 for further information
 Seq primer: M13 Reverse.

FEATURES

source
 1. 509
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RGIC011"
 /clone_id="Normalized rat embryo", Bento Soares"
 /dev_stage="embryo 8, 12, 18 dpc"
 /note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
 BASE COUNT 116 a 167 c 118 g 107 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 122; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaactgcttga 15
 |||
 Db 443 GGATCTGCTGTTGA 429

RESULT 27
 AA818607 528 bp mRNA EST 03-JUL-1999
 LOCUS UI-R-A0-bc-h-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
 DEFINITION UI-R-A0-bc-h-05-0-UI.3, similar to db|096683|MMU96683 Mus musculus
 immunoglobulin-like receptor PIRAZ (6M23) mRNA, partial cds, mRNA
 sequence.

ACCESSION AA818607 GI:4228398
 VERSION AA818607
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 TITLE Rattus.
 1 (bases 1 to 528)

JOURNAL
 MEDLINE
 COMMENT
 On Feb 17, 1998 this sequence version replaced gi:2889346.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9365
 Email: msoares@blue.weeg.uiowa.edu

FEATURES
 source
 1. 528
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-bc-h-05-0-UI"
 /clone_id="UI-R-A0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung, brain
 , liver, kidney, heart, spleen, ovary, and muscle. The tag
 is a string of 3-5 nucleotides present between the Not I
 site and the oligo-dT track which allows identification of
 the library of origin of a clone within the mixture."
 BASE COUNT 138 a 131 c 101 g 158 t
 ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaac 18
 |||
 Db 353 TCTGCTGTTGAAGC 339

RESULT 28
 TA375E11P/C 532 bp DNA GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 375e11, forward sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AL495591 GI:11873033
 VERSION AL495591.1
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.

REFERENCE
 AUTHORS Hall, N., Bowman, S., Leonard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajadream, M.A. and Barrell, B.G.
 TITLE Direct Submission

JOURNAL
 COMMENT
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 nhlee@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).

Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/T_brucei/.
 Location/Qualifiers
 1. 532

FEATURES
 source
 1. 532
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="375e11"
 BASE COUNT 162 a 156 c 92 g 122 t
 ORIGIN

Query Match 60.0%; Score 15; DB 258; Length 532;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgctgttgaagc 19
 |||
 Db 310 CTGCTGTTGAAGC 296

RESULT 29
AA790463/c 542 bp mRNA EST 06-FEB-1998
LOCUS
DEFINITION
AA790463.1 Soares_mammary_gland_NBMG Mus musculus cDNA clone
IMAGE:1230982.5, similar to SW:RAP3_HUMAN P17964 RAS-RELATED
PROTEIN RAP-2B. ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AA790463.1 GI:2850583
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 542)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepcie, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

COMMENT
JOURNAL
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MG1:656574
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 492.

FEATURES
source
Location/Qualifiers
1..542
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1230982"
/clone_1b="Soares_mammary_gland_NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland. Vector: pT773D-Pac (Pharmacia
) with a modified polylinker. Site_1: Not I. Site_2: Eco
RI. 1st strand cDNA was primed with a Not I - oligo(dT)
primer 15'.
TGTTCACATCTGAGTGGAGCGCGCGCAATGTTTTTTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), sites of the modified pT773 vector.
the Not I and Eco RI sites of the modified pT773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT
ORIGIN
127 a 178 c 125 g 112 t

Query Match 60.0%; Score 15; DB 12; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcgcgtttga 15
|||||
Db 486 GCATCTGCTTTTGA 472

RESULT 30
TA162C10P/c 559 bp DNA GSS 13-DEC-2000
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 162c10, forward sequence,
genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL472440
AL472440.1 GI:11837892
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 559)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA. E-mail: barrellesanger.ac.uk and
nhlesanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/projects/T-brucei/>.

FEATURES
source
Location/Qualifiers
1..559
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="162c10"

BASE COUNT
ORIGIN
198 a 154 c 92 g 115 t

Query Match 60.0%; Score 15; DB 258; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgcctgttgaacg 19
|||||
Db 188 CTCGCTGTGACGCG 174

RESULT 31
FR0036852 594 bp DNA GSS 22-OCT-1999
LOCUS
DEFINITION
Fugu rubripes GSS sequence, clone 039N10aE9, genomic survey
sequence.
AL124355
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FR0036852 594 bp DNA GSS 22-OCT-1999
Fugu rubripes GSS sequence, clone 039N10aE9, genomic survey
sequence.
AL124355
AL124355.1 GI:6105970
GSS; genome survey sequence.
Takifugu rubripes.
Takifugu rubripes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 594)
Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K.,
Umranta, Y., Williams, G. and Brenner, S.
Direct Submission
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
biohelp@hmp.mrc.ac.uk
Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic

FEATURES
 source
 Location/Qualifiers
 1. 594
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 039N10"
 /clone="039N10aP9"
 BASE COUNT 138 a 135 c 144 g 157 t 20 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 15; DB 222; Length 594;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaag 17
 |||||
 Db 475 ATCTGCTGTTGAAG 489

RESULT 32
 FR0035666 598 bp DNA GSS 22-OCT-1999
 LOCUS
 DEFINITION Fugu rubripes GSS sequence, clone 019B02CC2, genomic survey
 sequence.
 ACCESSION AL123183
 VERSION AL123183.1 GI:6104798
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 598)
 Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
 Umranta,Y., Williams,G. and Brenner,S.
 Direct Submission
 Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridge, CB10 1SB, UK Email:
 biohelp@hmp.mrc.ac.uk
 Vector: phagescript II KS
 V_type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source
 Location/Qualifiers
 1. 598
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 019B02"
 /clone="019B02CC2"
 BASE COUNT 120 a 147 c 153 g 165 t 13 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 15; DB 222; Length 598;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaag 17
 |||||
 Db 116 ATCTGCTGTTGAAG 130

RESULT 33
 FR0044579 600 bp DNA GSS 22-OCT-1999
 LOCUS
 DEFINITION Fugu rubripes GSS sequence, clone 192G14AF9, genomic survey
 sequence.
 ACCESSION AL132071
 VERSION AL132071.1 GI:6114017

KEYWORDS
 SOURCE
 ORGANISM
 GSS: genome survey sequence.
 Takifugu rubripes.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridge, CB10 1SB, UK Email:
 biohelp@hmp.mrc.ac.uk
 Vector: phagescript II KS
 V_type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source
 Location/Qualifiers
 1. 600
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 192G14"
 /clone="192G14AF9"
 BASE COUNT 123 a 152 c 153 g 170 t 2 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 15; DB 222; Length 600;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaag 17
 |||||
 Db 145 ATCTGCTGTTGAAG 159

RESULT 34
 FR0044632 611 bp DNA GSS 22-OCT-1999
 LOCUS
 DEFINITION Fugu rubripes GSS sequence, clone 192G14BC7, genomic survey
 sequence.
 ACCESSION AL132124
 VERSION AL132124.1 GI:6114070
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 611)
 Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
 Umranta,Y., Williams,G. and Brenner,S.
 Direct Submission
 Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
 Centre, Hinxton, Cambridge, CB10 1SB, UK Email:
 biohelp@hmp.mrc.ac.uk
 Vector: phagescript II KS
 V_type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES
 source
 Location/Qualifiers
 1. 611
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 192G14"
 /clone="192G14BC7"
 BASE COUNT 158 a 154 c 150 g 122 t 27 others

ORIGIN

Query Match 60.0%; Score 15; DB 222; Length 611;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttgaag 17
 |||||||
 DB 450 ATCTGCTGTTCGAG 436

RESULT 35
 BG306774 621 bp mRNA EST 22-FEB-2001
 LOCUS BG306774/c
 DEFINITION f06b08.y1 zebrafish adult retina cDNA Danio rerio CDNA clone
 4145535 5', similar to TR:Q920X1 Q920X1 APOPTOSIS-INDUCING FACTOR
 AIF.; mRNA sequence.

ACCESSION BG306774 GI:13104301
 VERSION BG306774.1
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio.

REFERENCE
 AUTHORS Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Actinopterygii: Neopterygii: Teleostei: Euteleostei: Ostariophysi:
 Cypriniformes: Cyprinidae: Rasbora: Danio.
 1 (bases 1 to 621)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 S., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.

TITLE Mashu zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: zbratfish@wustl.edu
 Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
 Washington University Genome Sequencing Center Clone distribution:
 RessourcenzentrumPrimatendank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: '3' ET from AmerSham
 High quality sequence stop: 420.

FEATURES
 source Location/Qualifiers

1..621
 /organism="Danio rerio"
 /strain="wild-type"
 /db_xref="taxon:7955"
 /clone="4145535"
 /clone_lib="zebrafish adult retina cDNA"
 /sex="mixed"
 /dev_stage="1-2 years"
 /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
 /note="Vector: Lambda ZAP II (pBluescript SK-); Site: 1:
 EcoRI; Site: 2: SalI; This Zebrafish library was
 constructed by Dr. Susan E. Brockerhoff (email:
 sbrocker@u.washington.edu) RZPD library number: 760"
 BASE COUNT 157 a 141 c 164 g 154 t 5 others
 ORIGIN

Query Match 60.0%; Score 15; DB 175; Length 621;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgaagcag 22
 |||||||
 DB 180 CTGTTGAAGCCAG 166

RESULT 36

BG017095 627 bp mRNA EST 24-JAN-2001
 LOCUS BG017095/c
 DEFINITION df68g03.x1 Xenopus laevis unfertilized egg cDNA library Xenopus
 laevis cDNA clone XENOPUS_SOURCE_ID:xlneeg005n05 3', mRNA sequence.

ACCESSION BG017095 GI:12470841
 VERSION BG017095.1
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE
 AUTHORS Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae:
 Xenopodinae: Xenopus.
 1 (bases 1 to 627)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
 Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
 B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 Mashu Xenopus EST project, 1999
 Unpublished (1999)
 Title: Other ESTs: df68g03.y1
 Contact: Sandy Clifton, Ph.D.
 Mashu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.edu
 Library constructed by Bruce Blumberg
 Library normalized by Jihwan Song
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clone distribution information for
 this library can be found through Research Genetics, visit their
 web page at: <http://www.resgen.com/> please reference the id listed
 below when ordering this clone: Source lab clone id - xlneeg005n05
 Seq primer: -40UP from Gldco
 High quality sequence stop: 468.

FEATURES
 source Location/Qualifiers

1..627
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="XENOPUS_SOURCE_ID:xlneeg005n05"
 /clone_lib="Xenopus laevis unfertilized egg cDNA library"
 /tissue_type="unfertilized egg"
 /lab_host="Top-10 F"
 /note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
 XhoI; cDNA was prepared from 2ug of poly A+ RNA. XR
 EcoRI-XhoI cut cDNA was then ligated into Unilap XR
 (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
 end. SS-library phagemids were prepared by mass excision
 from the original library and normalized by hybridization
 to biotinylated driver (prepared from the same library by
 PCR) to Cot-omega of 11. After removal of hybrids and
 excess driver by streptavidin sepharose chromatography,
 the ss-phagemids were made double stranded and
 electroporated into Top-10 F'. Original library
 constructed by Bruce Blumberg (Blumberg et al., 1991
 Science 253, 194-196; Hawley et al., 1995, Genes Dev. 9,
 2923-2935). Note: This is a Xenopus Gene Collection (XGC)
 library."
 BASE COUNT 146 a 131 c 143 g 206 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 172; Length 627;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tttaagcagcag 25
 |||||||
 DB 556 TTGAAGCCAGCAG 542

RESULT 37
A0754510 632 bp DNA GSS 27-JUL-1999
LOCUS HS_5434.AL.G05.SP66.RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone plate=1010 Col-9 Row-M, DNA sequence.
ACCESSION A0754510
VERSION A0754510.1 GI:5618726
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.hsc.washington.edu/plate/1010_row:M_column:9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 632.
Location/Qualifiers
1..632
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1010 Col=9 Row=M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 156 a 143 c 126 g 190 t 17 others
ORIGIN

Query Match 60.0%; Score 15; DB 233; Length 632;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gctcttgaagcgca 21
|||||
Db 308 GCTGTTGAAGCGCA 322
RESULT 38
BG546580
LOCUS BG546580 640 bp mRNA EST 04-APR-2001
DEFINITION 602574435f1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702473 5',
RNA sequence.
ACCESSION BG546580
VERSION BG546580.1 GI:13545245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens .

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 640)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLCM1539 row: f column: 10
High quality sequence stop: 640.
Location/Qualifiers
1..640
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4702473"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site.1:
SfiI (ggcgccgctggcc); Site.2: SfiI (ggccatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATTAATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

FEATURES
source
BASE COUNT 170 a 131 c 155 g 184 t
ORIGIN

Query Match 60.0%; Score 15; DB 155; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 atctctgttgaag 17
|||||
Db 600 ATCTGCTTTGAAG 614

RESULT 39
BE597936
LOCUS BE597936 645 bp mRNA EST 18-AUG-2000
DEFINITION P11_67_C08_g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
RNA sequence.
ACCESSION BE597936
VERSION BE597936.1 GI:9853009
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 645)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: POLYTMix
High quality sequence start: 3
High quality sequence stop: 635
POLYA=No.

FEATURES

SOURCE

Location/Qualifiers
1.645
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate PRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 179 a 143 c 170 g 153 t
ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 645;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgtttgaagc 19
|||||
Db 627 CTGCTGTTGAAGCG 641

RESULT 40
BG085090/c 649 bp mRNA EST 26-JAN-2001
LOCUS BG085090-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H3108E05 5', mRNA sequence.
ACCESSION BG085090
VERSION BG085090.1 GI:12567654
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Karaul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
1 (bases 1 to 649)
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Other_ESTS: H3108E05-3
COMMENT Unpublished (2001)
Contact: George J. Karaul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnaelg@sun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3108 ROW: E column: 05
Seq primer: -21M13 Reverse
High quality sequence stop: 649
POLYA=No.

FEATURES
SOURCE 1.649
Location/Qualifiers
/organism="Mus musculus"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3108E05"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source cDNA libraries are cloned unidirectionally with 01190(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 157 a 182 c 177 g 133 t
ORIGIN

Query Match 60.0%; Score 15; DB 173; Length 649;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttga 15
|||||
Db 315 GGATCTGCTGTTGA 301

RESULT 41
AM633493 663 bp mRNA EST 03-APR-2000
LOCUS AM633493
DEFINITION laevis cDNA clone pX0008D10 5', mRNA sequence.
ACCESSION AM633493
VERSION AM633493.1 GI:7390574
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae; Xenopodinae; Xenopus.
AUTHORS 1 (bases 1 to 663)
Blackburn, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G., Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
J.W., Bonaldo, M.F., and Soares, M.B.
TITLE The NIH's Xenopus Maternal EST Project
JOURNAL Unpublished (2000)
COMMENT Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709, USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial Parkway, Huntsville, AL 35891
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC).

Health Intramural Sequencing Center (NISC).

PCR Primers
 FORWARD: TGTAAACGACGCGCAGT
 BACKWARD: CAGGAACAGCTATGACC
 Plate: 0008 Row: D Column: 10
 Seq primer: T7 primer.
 Location/Qualifiers
 1. 663

FEATURES
 source
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone_lib="PBX0008D10"
 /clone_lib="Blackshear/Soares normalized Xenopus egg library"
 /sex="female"
 /tissue_type="unfertilized egg"
 /cell_type="unfertilized egg"
 /dev_stage="unfertilized egg"
 /lab_host="DH10B"
 /note="Vector: pT73-Pac; Site:1: EcoRI; Site:2: NotI; Polya-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri8 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10⁵ recombinants, with average insert sizes of 1-1.5 kb."
 BASE COUNT 211 a 172 c 139 g 140 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 118; Length 663;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ttctgaagcgcagcaag 25
 |||
 Db 237 TTTGAAGCGCAGCAG 251

RESULT 42
 BE820949/c 681 bp mRNA EST 21-SEP-2000
 LOCUS GM700013A20C12 Gm-r1070 Glycine max cDNA clone Gm-r1070-5112 3',
 DEFINITION mRNA sequence.
 VERSION BE820949
 KEYWORDS BE820949.1 GI:10253183
 SOURCE EST.
 ORGANISM soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 681)
 Vockin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvelli, V., Expanding, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H. A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)
 Other-ESTs: AW432846 corresponding to Gm-cl016-6609 (5')
 Contact: Vockin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vockin@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 pr (314) 427-3222 FAX:(888) 919-3524 or (314)

427-3124 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers
 1. 661

FEATURES
 source
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-5112"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries were used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html. Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."
 BASE COUNT 190 a 142 c 132 g 209 t 8 others
 ORIGIN

Query Match 60.0%; Score 15; DB 140; Length 681;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaa 16
 |||
 Db 490 GATCTGCTGTTTGA 476

RESULT 43
 AM128741/c 748 bp mRNA EST 25-OCT-1999
 LOCUS fe36h03.x1 zebrafish Washu MPMG EST Danio rerio cDNA 5', similar to
 DEFINITION TR:0920X1 Q920X1 APOPTOSIS-INDUCING FACTOR AIF.; mRNA sequence.
 VERSION AM128741
 KEYWORDS AM128741.1 GI:6116645
 SOURCE EST.
 ORGANISM zebrafish.
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbortinae; Danio.
 1 (bases 1 to 748)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kneba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schur, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished (1998)
 Other-ESTs: fe36h03.x1
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark, CDNA Library Arrayed by:
 Matthew Clark, DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 Resourcenet/centuriprimatebank, Berlin, Germany (web address:
www.rzpd.de)

FEATURES

source

Seq primer: T3 ET from Amersham
 High quality sequence stop: 487.
 Location/Qualifiers
 1. 748
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="Zebrafish Washu MPMG EST"
 /sex="mixed"
 /tissue.type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
 strand cDNA was primed with a Not I oligo(dT)15 primer
 15' PGACGATGCTTACATCGCAGCGCCGCTTTTCTTTTCTTTT3';
 double-stranded cDNA was ligated into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 183 a 186 c 197 g 181 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 112; Length 748;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgagcgagcag 22
 |||||
 Db 415 CCGTTTGAAGCGCAG 401

RESULT 44
 CDS02060 752 bp DNA GSS 14-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence pUC-Or1 end of clone
 DEFINITION 152M04 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 AL206315.1 GI:7865134
 GSS: genome survey sequence.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 752)
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the

REFERENCE
 AUTHORS
 TITLE

JOURNAL

Unpublished
 2 (bases 1 to 752)
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Mincker,P., Brotlier,P., Queller,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

TITLE

JOURNAL

Unpublished
 3 (bases 1 to 752)
 Genoscope.

REFERENCE

Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

COMMENT

FEATURES

source

Location/Qualifiers
 1. 752
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="152M04"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG152DG02SP1-end :
 pUC-Or1"

BASE COUNT 138 a 199 c 238 g 163 t 14 others
 ORIGIN

Query Match 60.0%; Score 15; DB 220; Length 752;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ttgagcgagcagcag 25
 |||||
 Db 698 TTTGAAGCGCAGCAG 684

RESULT 45
 BC428569 846 bp mRNA EST 14-MAR-2001
 LOCUS 602501093F1 NIH_MGC-75 Homo sapiens cDNA clone IMAGE:4614647 5',
 DEFINITION mRNA sequence.
 BC428569
 BC428569.1 GI:13335075
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LINC1366 row: j column: 24
 High quality sequence stop: 678.
 Location/Qualifiers

FEATURES

source

1. 846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:4614647"
 /clone_lib="NIH_MGC_75"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:

SfiI (ggcgccctcgcc); Site 2: SfiI (ggccattatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGCGCGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 235 a 214 g 187 t

ORIGIN

Query Match 60.0%; Score 15; DB 153; Length 846;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgtttgaa 16
 |||||
 Db 732 GATCTGCTGTTTGAA 718

Search completed: October 9, 2001, 15:15:28
 Job time: 13650 sec


```

9      25      100.0      5872      9      AR101809      Sequence
10     25      100.0      5872      9      AR101810      sequence
11     25      100.0      11022      1      AE000180      AE000180 Escherich
12     25      100.0      13501      1      AE005258      AE005258 Escherich
13     25      100.0      297816      2      AP002553      AP002553 Escherich
14     22      88.0      1121      10      E00893      E00893 Genomic DNA
15     20      80.0      2923      3      MAV250020      MAV250020 Mycobacte
16     19      76.0      3603      3      PSEBHP1A      L11582 Pseudomonas
17     19      76.0      44201      3      MTCY180      297193 Mycobacteri
18     18      72.0      37500      73      AC068068      Leishmani
19     18      72.0      38000      61      AC010079      Leishmani
20     18      72.0      42000      77      AC087836      Leishmani
21     17      68.0      12891      1      AE004192      AE004192 Vibrio ch
22     16      64.0      4674      89      AK024397      AK024397 Homo sapi
23     16      64.0      10290      1      AE005685      AE005685 Caulobact
24     16      64.0      13350      3      MTCV031      AL021926 Mycobacte
25     16      64.0      23740      3      MTCV026      AL022076 Mycobacte
26     16      64.0      34902      3      POL245436      AL133503 Pseudomon
27     16      64.0      49587      6      DMH48C10      AL133503 Drosophill
28     16      64.0      60573      65      ATAC009895      AC018267 Drosophill
29     16      64.0      82289      13      AC009895      Arabidops
30     16      64.0      115955      62      AC011732      AC011732 Homo sapi
31     16      64.0      177380      85      AC002457      AC002457 Homo sapi
32     16      64.0      180700      76      AC084136      AC084136 Homo sapi
33     16      64.0      289090      4      AE003424      AE003424 Drosophill
34     15      60.0      893      53      CENS06ZHL      AL147200 Anopheles
35     15      60.0      978      53      CENS06ZHL      AL422335 T3 end of
36     15      60.0      3069      13      AF219383      AF219383 Drosophill
37     15      60.0      3369      13      AF274670      AF274670 Telraden
38     15      60.0      3369      13      AF274671      AF274671 Trochoden
39     15      60.0      13916      3      MTCY9CA      AF065159 Bradyrhiz
40     15      60.0      35209      1      AF065159      AC009603 Leishmani
41     15      60.0      35573      4      AC009603      AC013257 Leishmani
42     15      60.0      41193      4      AC013257      AC010753 Drosophill
43     15      60.0      44545      61      AC010753      AC014684 Drosophill
44     15      60.0      55307      63      AC014684      AC014903 Drosophill
45     15      60.0      73691      63      AC014903

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source
1. 1041
Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggcgcagcagtgatgcgcagcatt 25
|||||
Db 79 GCGCAGCAGTGATGCCAGCATT 103

```

```

RESULT 2

```

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source
1. 1041
Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggcgcagcagtgatgcgcagcatt 25
|||||
Db 79 GCGCAGCAGTGATGCCAGCATT 103

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION Biot gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
source
1. 1084
Location/Qualifiers
/organism="Escherichia coli"
/db_xref="taxon:562"
24. 1064
/gene="BIOB"
24. 1064
/gene="BIOB"
/codon_start=1
/translation="MAHRRKRLTSYTELFEKPLDLLEPAQVHROHFFDPROYVST
LISIKGACPEDCKTCYQSSSRKRTGLEAEKRLMEVQVLESARAKAAGSTRFGAGAAW
KNPHERDMPLBDMVGVKAWGLEACMTGLTSEQAORLANAGLDYNNHLDLSPES
YGNITITTRYOERLDTLEKVRDAGIKVCSGIGVIGETVKDRAGLLDOLANLPPPPS
VPINMLVKVGTPLADNDVDVDAFDEIRTAIVARIIMPTSYVRLSAGREONBOQAAC
FMAGANSIFYGCKLITTPNPRBEDKDLQLEFRKLGILNPQQTAVIAGNBQQRLBQALMT
PVDDEVYNNAAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatgcgcagcatt 25
|||||

```

CDS

CDS

CDS

E C O L

20

CESSIT

URCE
ORGAN

FERENC
AUTHOR
TITLE
JOURNAL

gene
3049. .4203
/gene="bioC"
3049. .4203
/gene="bioC"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"
/translation="MSMOEKTNALDARRAADAALRRPVPAOGAGRMVLVADDDRLNF
/SMDYIGLSHHNPQITRANOGCKRPFQSGSGSHVGYCVHQALEELAEMLSTSA
LDTISGPAANOAVIAMMAKEDRIAADRLSHALESALSPSOARRAHNDYGNR
LMSPGGOOMVTVTEGVSMGDSAPRLAETIOYQNHMLMWDDAHGTGYGOGHG
CSLQVAKRDLVTFEGVFGVSGAVALVCSYVADLLIDPARHILYISNMPAAQAL
RASIAIVRSRSGIARRERKIALILTRFPRGVQDLPTLADSCSAIOPITLVGNSAIDL
AEKIRFOOGCVWEVAIRPVPVPAGTARLRLITLAAHEMDIDRLLEVLRNG"

gene
4190. .4945
/gene="bioC"
4190. .4945
/gene="bioC"

BASE COUNT	1363 a	1554 c	1631 g	1245 t
ORIGIN	4626 bp upstream of HpaI site; 18 min on K-12 map			

Query Match	100.0%	Score 25	DB 2	Length 5793
Best Local Similarity	100.0%	Pred. No. 0.0019		
Matches	25	Conservative 0	Mismatches	Indels 0
				Gaps 0
QY	1	gagcagcagtgatccgcagcatt	25	
DB	2090	gGcAGcAGtgATccCGcAGcATT	2114	

RESULT	5			
A38246				
LOCUS	A38246	5872 bp	DNA	PAT
DEFINITION	Sequence 1 from Patent WO9408023.			
ACCESSION	A38246			
VERSION	A38246.1	GI:2294844		
KEYWORDS				
SOURCE	Escherichia coli.			
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.			
REFERENCE	1 (bases 1 to 5872)			
AUTHORS	Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.			
TITLE	BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN			
JOURNAL	Patent: WO 9408023-A 1 14-APR-1994;			

COMMENT LONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 location/Qualifiers
 1. 5872

FEATURES
 source
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PHO30A-15/9"
 1. .96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23. .28
 45. .49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105. .119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117. .1157
 /gene="BIOB"
 117. .1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHPRKWTLSQVTELEFKPLDLLFEAQVHRQHPDPOVST
 LLSITGACPEDECKYCPQSSRYKGTGLEAEERLMEVQVLESARAKAGSTRFGMAW
 KNHREDDYERLEQWQVQKAMGLACMTLGLTSSQARLANGLDYNINLTPSEF
 IGNITTRPYERLDTLEKVDAGIKVCSGIVGLGETVRDRAGLILQIANLPTPES
 VPIMLVYKGTPLADNDVDAPDFRTIIVARIIMPTSYRISAGREONEDTQMC
 FMAGNSIFYCKSLTTPNPREDKIDLFKRLGIPQQTAVLAGDNBQQRLEQALMT
 PDTDEYVNAAL"
 2284. .2297
 /standard_name="BIOC RBS"
 2295. .3050
 /gene="BIOC"
 2295. .3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVNRQALIAAAGRAAAHYEQHADLQROSADALIAMLPORKY
 HYLDAGCGPEGMSSRHWRERHAQVATLALDSPMLVQARQDADHYLAGDIEPLATA
 TFDLAWSNLAWQCNLSTALRELYVVRPKGVATTLVQSLPELIGQANQVDERP
 HANRFLPDEPLEISLNGVYOHIIQITLMPDALSAMSLKIGTATHLHREGDRILL
 TRSOLRQLAWAQOQGRYPLTYHLFGLVIAAE"
 3742. .3752
 /standard_name="BIOA RBS"
 3750. .5039
 /gene="BIOA"
 3750. .5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OKONONANATE AMINOTRANSF.""
 /RC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5

/evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDLAFDORHIMHPTSMSTPLPYPVVSAEGCELILSDGR
 LVMGSSWMAAIIHGNHNPOLNMAKSOIDAMSHVMEGGITTAPELCECRKIYAHTPOP
 LEVFLADSGSVAVEAMKALQYQAKGAEARQRLTTRNGHGTDFGMSVCDPDS
 MSLMKGYLEPENLFAPQSRMDGEWDERDNGFRRLMAAHHEIAVAVIEPIVQAG
 GMRVHPENMLKRIKRICDREGILLIADETATGFRGKLFACERHAELAPDILCGKAL
 TGGTMTLSATLTTRVEAETISNGEAGCEHGGTFMGNPLACAAANASIALIESGDMOQ
 QVADIEVQLREQLAPARDAEMVADVRLGALIGVETTPVNMALOKFVQGVNTP
 FGLIYIMPEYIILPQOLRLTAANRAVQDETFCQ"
 5088. .5100
 /standard_name="ORF1 RBS"
 5098. .5574
 /gene="ORF1"
 5098. .5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGDKLPHRHYNGMGYGDNI SPHLANDVPACTK
 SEVVCYDPAIDPTGSGMHWVYVNLDPDTRVLPDGRSGGLVAMPDGLQRTDPGRTG
 YDGAAPKRGERTHRYLFTYHALDTRIDVDGASGAMGVNHFHSLASITVAFS"
 5583. .5605
 5583. .5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 25; DB 9; Length 5872;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtcctcgcgcagcatt 25
 |||||||||||||||||||||||||
 Db 195 gccgcagcaggtcctcgcgcagcatt 219

RESULT 6
 A38251 5872 bp DNA PAT 05-MAR-1997
 LOCUS
 DEFINITION Sequence 6 from Patent W09408023.
 ACCESSION A38251
 VERSION A38251.1 GI:2294849
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 6 14-APR-1994;

COMMENT LONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 location/Qualifiers
 1. 5872
 /organism="Escherichia coli"

```

RBS      /strain="DSM498"
          /db_xref="taxon:562"
          /clone="PB030A15-9"
          1141..1156
          /standard_name="BIOF RBS"
          1154..2308
          /gene="BIOF"
          1154..2308
          /gene="BIOF"
          /standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
          /EC_number="2.3.1.47"
          /codon_start=1
          /transl_table=11
          /number=2
          /evidence=experimental
          /product="KAPA SYNTHASE"
          /protein_id="CA02329.1"
          /db_xref="GI:2294850"
          /translation="MSWQKINAAIDARRADALRRRYPAAGAGRWLVADROYLN
          SSNDYGLSHHPQIIRAMQGAOEGIGSGGSHVSGYSVHQALEELAEWLGYSRA
          LFIISGPAANOAVIAAMAKEDRIADRLSHASLLEASLSPQLRFAHNDVTHLAR
          LASRPGCOMVTEGVSMDSAPLAEIOOVTOOHNGMLWVDAHGTVAGEQGRG
          SCWLOKYPRLVTVTGKGFSGVGAIVLCSTVADYLOFARHLIYSTMPAQOAL
          RASIAVIRSDGAREKUALITTRACVODPLTADSCAIOPLIVGMSRAIOL
          AEKIROGQCMVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
          3030..3045
          /standard_name="BIOB RBS"
          3043..3753
          /gene="BIOB"
          3043..3753
          /gene="BIOB"
          /standard_name="DETHIOBIOTIN SYNTHASE"
          /EC_number="6.3.3.3"
          /codon_start=1
          /transl_table=11
          /number=4
          /evidence=experimental
          /product="DTB SYNTHASE"
          /protein_id="CA02330.1"
          /db_xref="GI:2294851"
          /translation="MSKRYEVTGDTFEVGTQVASCALLQAAAGYRTAGYKPVASGS
          EKTPEGLRNSDALIORNSSLOLDYATNPYTFAPETSPHIIISAQGRPEESLMSAG
          LRALBQADWVLVSGAGCMPTPLSDFTFADWTOBOLPVLIVGVKICINIAMLTA
          QVIOHAGLTLGAWANDVTPPGKRAHEVMTTLTRMIPAPLLGLIPMLAEPEMAATGK
          YINLAIVDASTLIGTSRL"
          1552 c 1695 g 1307 t

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatcgccagcatt 25
    |||
Db 195 ggcgcagcagtgatcgccagcatt 219
    |||

RESULT 7
A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;

```

```

FEATURES             LOCATION
source               1..5872
                    /organism="Escherichia coli"
                    /strain="DSM498"
                    /db_xref="taxon:562"
                    /clone="PB030A-15/9"
promoter             1..96
                    /function="PROMOTER PTAC"
                    /evidence=experimental
-35_signal           23..28
-10_signal           45..49
                    /standard_name="PROMOTER PTAC"
                    /evidence=experimental
RBS                 105..119
                    /standard_name="BIOB RBS NO.9"
                    /evidence=experimental
gene                117..1157
                    /gene="BIOB"
                    /db_xref="GI:6741863"
                    /transl_table=11
                    /codon_start=1
                    /evidence=experimental
                    /product="BIOTIN SYNTHASE"
                    /protein_id="CAB69590.1"
                    /db_xref="GI:6741863"
                    /translation="MAHRPWTLTQVTELEFEKPLDLLFEAQOVRHOFDPKOVST
                    LLSKTGACPEDCKYCQSSRYKTELEARLMEVEQVLEESAKRAKAAGSTRGCAAM
                    KNPERMDPYLEQWGVKAMGLEACMTIGTSESOABRLAANGIDYVNHNLDTSEF
                    YGNIITRTYQERLDLTKYVDAGIKVSGGIVGLEYTKYKDAIGLLQLANLPPEES
                    VPIMLTKVKGTEPLADNDVDADPFI RTIAVARIMPTSVIRLSGREGHNEQYAMC
                    FMAGANSIFGCKLITTPNEDEKDLQLEFRKILNPOQAVLAGNEQQOQRLQALMT
                    PTDEYVNAAL"
                    2284..2297
                    /standard_name="BIOC RBS"
                    2295..3050
                    /gene="BIOC"
                    2295..3050
                    /gene="BIOC"
                    /function="INVOLVED IN PIMETOYL-COA SYNTHESIS"
                    /codon_start=1
                    /transl_table=11
                    /product="PROTEIN"
                    /protein_id="CAB69591.1"
                    /db_xref="GI:6741864"
                    /translation="MATVKNQAIANAFAFGRAAHYEQADLQROSADALLAMLPORYT
                    HYLDAGCGPGMSRHHRRRAQVATLDSPPMIVQARQKADADHYLAGDIESLPLATA
                    TFDLAMSILAYOKGNCSTALRELXYRVRRGVYAFPTLVQSLPELHQAQMVDERP
                    HANRFLPDETEIQSLNGVYHGHIIQPTITMPPDALSMSRLKIGATHHEGSDPRL
                    TRSQLRQLQALMPQOQGRYPLTYHLFGVLAKE"
                    3742..3752
                    /standard_name="BIOA RBS"
                    3750..5039
                    /gene="BIOA"
                    3750..5039
                    /gene="BIOA"
                    /standard_name="S-ADENOSYL-L-METHIONINE-8-AMINO-7-
                    OXONONANATE AMINOTRANSF."
                    /EC_number="2.6.1.62"
                    /codon_start=1
                    /transl_table=11
                    /evidence=experimental
                    /product="DAPA SYNTHASE"
                    /protein_id="CAB69592.1"
                    /db_xref="GI:6741865"
                    /translation="MTTDLAPDRHHIWHPTYSMTSPULYVPVVSABCEILLSDGR
                    LVDGSSMWAALIHGVNHPOLNAAKMSQIDAMSHVFGGITHAPALEICRKYVANTPOP
                    LFCVPLADSGVAIVFAKMALOYQAQGEARQFLTFRNQYHGDTEGAMVCDPDS
                    MSLKMGCIYIPENLAFAPQSRMDGEMDERDVGAFARLAAHREHIAVYIPIYOGAG
                    GMRWYHPEWLKRIKTCDBREGILLIADRLATGEGFTGKLPACENAEIADILCLGAKL
                    TGCTMTLSATLTTRVAETISNGEAGCTRHGPTPGNPLACANAASIALIESDMDQ
                    QVADIEVQLREQLAPARDAEMVADVAVIGALIGVETTHPVNMAALQKFEVQGWIRP

```

```

RBS      5088..5100      /standard_name="ORF1 RBS"
gene      5098..5574      /gene="ORF1"
CDS       5098..5574      /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="PROTEIN"
                        /protein_id="CAB69593.1"
                        /db_xref="GI:6741866"
stem_loop /translation="MKLISNDLRDCKLPKRVFNGMGYDGNISPLAMDVPAGTK
terminator YDGAAPKGTIMHYITFHALDIERIDVDEGASGAMVFNFHSLASATVAMS"
                        5383..5605
                        5583..5644
                        /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatgcgcagcatt 25
Db 195 GCCGACGAGTGATGCACGACGATT 219

RESULT 8
LOCUS      A93679      5872 bp      DNA
DEFINITION A93679      Sequence 6 from Patent EP0798384.
ACCESSION  A93679
VERSION    A93679.1  GI:6741867
KEYWORDS
SOURCE     Escherichia coli.
ORGANISM   Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O. and Brass, J.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
          LONZA AG (CH)
FEATURES
Source     1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="BIOP RBS"
            1154..2308
            /gene="BIOP"
            1154..2308
            /gene="BIOP"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAB69594.1"
            /db_xref="GI:6741868"
            /translation="MSMDKINAAIDARRADALRRRYPVAGAGRWLVADROYLNF
SSNDYIGLSHHFOLLIRAWOOGAEGFGISGGSGYGVVHQALEELFAMLELGYRA
LFTISFAANQAVIAAMAKEDRIADRLSHASLLEASTSPSOLRRFRANDVYHLAR

```

```

RBS      3043..3753      /standard_name="BIOD RBS"
gene      3043..3753      /gene="BIOD"
CDS       3043..3753      /standard_name="BIOD RBS"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="DTB SYNTHASE"
                        /protein_id="CAB69595.1"
                        /db_xref="GI:6741869"
                        /translation="MSKRYFVGTDEYGTIVASCALLOAKAGRYFAGYKPVASG
                        ERTPEGLRNSDALORNSLSOLDVATNPYTPAEPSPHITSQEBRPTESLVSAG
                        LRALBOADWVIVGAGAGMPTLSPFTFADWYVOBQPVILVGVKLGICINHAMLTA
                        OYIOHAGLTLACWANDVTPPKRAEYMTTLITRMIPAPLIGELPWLAEINPENAATGK
                        YINLAPVDASTLGTLSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatgcgcagcatt 25
Db 195 GCCGACGAGTGATGCACGACGATT 219

```

```

RESULT 9
LOCUS      AR101809      5872 bp      DNA
DEFINITION AR101809      Sequence 1 from patent US 6083712.
ACCESSION  AR101809
VERSION    AR101809.1  GI:12812607
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
          Location/Qualifiers
          1..5872
          /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatgcgcagcatt 25
Db 195 GCCGACGAGTGATGCACGACGATT 219

RESULT 10
LOCUS      AR101810      5872 bp      DNA
DEFINITION AR101810      Sequence 6 from patent US 6083712.
ACCESSION  AR101810
VERSION    AR101810.1  GI:12812608
KEYWORDS

```

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-Jul-2000;
FEATURES
Location/Qualifiers
source
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcgcagcagtgatgcgcagcatt 25
Db 195 GCGCAGCAGTGATGCCAGCAT 219
RESULT 11
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
DEFINITION genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS
SOURCE ORGANISM
Escherichia coli K12.
Escherichia coli K12.
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
Escherichia.
1 (bases 1 to 11022)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gabor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
9278503
2 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolleg@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolleg@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli genetic

Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible at
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products: all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source
Location/Qualifiers
1. 11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1. 130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147. 1430)
/gene="ybhC"
/note="b0772"
complement(147. 1430)
/gene="ybhC"
/function="orf; Not classified"
/note="f427: 98 pct identical to fragment YBHC_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=11
/product="putative peptidase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNFESVSRLLALAPGVTLTACSSPPDQRPDSOTAPETSSRPI
LSAEADNPDAOHYFASLTPGAAMNPSPTLTPAODPFVAGTGVHTTIGQAVD
AATIKRTKROYIAVMEGYOGGVYPAAPAGGTLTGKRPIDVKGISLDGKSPA
DMRHDVDPKGYMPGKPAWYATYDSQSKRSDSGVCSAVFSONNGLOLNTLENT
LGSDVDAGNHFAVALRTDGDQVOJNNVNLIGRQNTFPVNSGYNLEFNRPRPRLVT
NSYIEGVVDIVSGGAVVPDTEFRVNSSTQGEAVVFAPATLSNITTYGFLVNSGFN
AFGGVADLGSLDVAANTNGOVYTRDSAINEGFNAPKPAWADAVISNRPAGNTGSVD
DNDERQRLNDNTNRYRMEXYNNRCVGSKEAAK"
complement(1507. 1535)
/note="Factor Sigma70; predicted +1 start at 806574"
complement(1582. 2058)
/gene="ybhB"
/note="b0773"
complement(1582. 2058)
/gene="ybhB"
/function="orf; Unknown"
/note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRQDGLPHRHVFNMGKGDGDNISPHLADNDVPAGTK
STFVATVDADATGSGMHWVYVNIAPATRYLQFSGGLVAMPDGLQTRTDEGKSG
YGCAPPKGERTHYIFVHAADIERIDVDEKASAMGCFNHFHSLASASITLAFS"
complement(2108. 2124)
/note="central position to predicted promoter:85.5"
/bound_moiety="Rnas predicted site"
complement(2117. 3406)
/gene="b10A"
/note="b0774"
complement(2117. 3406)
/gene="b10A"
repeat_region
gene
CDS
protein_bind
gene
CDS

Query Match	100.0%	Score 25	DB 1	Length 11022
Best Local Similarity	100.0%	Pred. No. 0.0017		
Matches 25	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	gcgcagcaggtgcacgcacgcatt 25			
Db 3571	GCGCAGCAGTCATCGCCAGCATT 3595			
RESULT 12				
AE005258				
LOCUS	AE005258 13501 bp DNA BCT 21-MAR-2001			
DEFINITION	Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82 of 155.			
ACCESSION	AE005258 AE005174			
KEYWORDS	AE005258.1 GI:12513751			
SOURCE	Escherichia coli O157:H7 EDL933.			
ORGANISM	Escherichia coli O157:H7 EDL933			
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia			
AUTHORS	1 (bases 1 to 13501)			
	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.			
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7			
JOURNAL	Nature 409 (6819), 529-533 (2001)			
MEDLINE	21074935			
PUBMED	11206551			
REFERENCE	2 (bases 1 to 13501)			
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA			
FEATURES				
source	Location/Qualifiers			
	1..13501			
	/organism="Escherichia coli O157:H7 EDL933"			
	/strain="EDL933"			
	/serotype="O157:H7"			
	/db_xref="taxon:155864"			
	/note="enterohemorrhagic"			
	<1..7576			
misc_feature	/note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933k; includes one copy of the 13 bp direct repeat that flanks the prophage"			
	66..665			
gene	/gene="lomK"			
	/note="20981"			
	66..665			
CDS	/gene="lomK"			
	/function="putative membrane; other or unknown (phage or prophage related)"			
	/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g117532789[gb] AAFG5231.1[AF151091_2 (AF151091)] lom [prophage P-Elba]"			


```

/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MKVCAAILLSAATLVASVPANASEHQSTLSAGYLHASTDAPC
SDDLINGINVKRYEFTDPLGLTFSYNADEQKTHSDTRHEDYVRNRRMSVAG
PSAVNEMESAAMAGAVASRYSTFGDYFRTDNKRTHDVLGSDARYSNTSLAM
GAYOFNPTESAVDAYEXSGSGDMRTDGFIVGYKF"
719..2041
/gene="20982"
/feature="20982"
/function="putative structure; Structural component (phage
or prophage Related)"
/notes="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from GenPept 118 :
g1145854361gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVQISGVKDCAGKPIQNCITOLKARRNSYXVVNTVASE
NPDEAGRYSDVEYGOYVTLVEGFPSPHAGTISVIEDSQGTINDFLGAMTEDAK
PEALRRPEQWVEEAARHAEARKNAGEAETSARNAGISAKSAANAADVSADASE
SARQAESASAKSEASASSASASASASASASASASASASASASASASASASAS
KSTTERARSAESASASASASASASASASASASASASASASASASASASASAS
KQTPGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
GPPGPGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
TGLTFDPPKATOVGMFQHLQVRFEGDFWQDKGLDEVSGTGRTGE"
2004..2312
/gene="20984"
/feature="20984"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from GenPept 118 :
g1145854371gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIQAELENDMNLIRKLMSLGCCKHDDCENGSRSLTAQLR
LGPADLLESDENGILPEQDRVITQVAILDADKKQICCVVAPLQILRADGTWENIGMK"
2489..3469
/gene="20985"
/feature="20985"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 :
g1169603671gb|AF33527.11 (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPITRTFNSVNIYVSPQVTSFAGKEYELKVIDETPILF
OMFENPERKXDEVPYNTKQHPYLDNVTNAARIESDMIGIFVSGDFSVNQTAFS
KLRFENPEWMTIYREDVDSWDRKLSYDIHDIQFQRLRTEDKRDYELNLEKEIR
ETSKADSLISMYAKKRHHAWPFFPRNIALIKAGETRCYNTNRKHGISEGCTYLD
MDMTITGKLTGTVAPDGLSHVDRRNDVSNVENSATITVNSNHPALLEGISPMHSKYD
AHPYVDGLGKGVKKYFNTPLANYNHFCDFLEFNHPIIMNTSOTYTCSSW"
3530..4522
/gene="20986"

3530..4522
/gene="20986"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFESKAPAGYSAPAPNREANAYADYVLDIGKRIP
SAADLSNVEYSYRAVHDSRSLIDQHTVDNIGVTDALSRSTQFDVASTYGINNEK
VHIGCIKRYNEEYELNEESSYKIDDIQSLTCLNELIYDYGQDPIPEICEADENDEEY
VPSVYAPDQDSYEMPSQMOEGILHEIHHVTGSSDPSGDSNTEIGPTILARRVAQELG
MSVPDEKGYAFEREERAHILRLNLRLQAAHREHNERAFPERLGTISDRYASPDFT
EYSANSNIGTGTQDHDPGLAINDNLQDANQIDQYHAPYIFPFEGVDKHNQ"
5350..6231
/gene="20989"
/feature="20989"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPISVNIQCSMNSTFRNLTPSPNRVLSVYDAVAHSDNGAOV
KVNKRYRVAVATDNKFCVTRRESHGFTNLRLKMPKGEISRKLEWNLSPVSAAM
ERGIVSHNSRPDLPPVDYAPPELPSVYNLSPGVANTIGGAAYVIEDAEDATYLA
FTTSQSNSEEVTSSEVRCENQYVYGASAEKLYGNNGDITGRMDKINGESLNLISLP
AERHATDMFRLKOKGILFVDITETVLYDRAKNFNPIDISSYVNSDRSSEQIMO
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
/feature="20990"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPSTNLVLIHQSSRSSMSDIDIESLVKASSQWIKNNPOLRF
OCTHNTYQDTEALDKIGSTETGRVILNATESISRLKSEYVYIHLNSSLRGVMAHD
IDAENHNGTSDPFCNLNAYEPCGBCISYVDHATVHEIHLNPHNLNGRLVYES
SRASQKYSPLLEEARVGLGAFSEVSESNKFHEIIGMPTRTSTPKXSLIHDDNT
VSLGQOVRHLHPIL"
7668..8144
/gene="20992"
/feature="20992"
/function="orf; Other or unknown (phage or Prophage
Related)"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLPKRHVFNMGVYDGDNIISPLHAMDVPAGTK
SFVYCTYPPDPTSGGMMHWYVVLPAADTVLPQGGSGGLVAPDDGLQTRDFGKTK
YDGAAPFGETHRTITVHALDVRIDYDEGASGAMGFVNHHSLSASATIMFS"
complement(8203..9492)
/gene="20993"
/notes="20993"
complement(8203..9492)

```

Query Match 100.0%, Score 25; DB 1; Length 13501;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcagatgcacgcagcatt 25
 ||||||||||||||||||||
 Db 9657 GCGACGACGTCGTCGCCGACATT 9681

RESULT 13
 APO02553 297816 bp DNA BCT 07-MAR-2001
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
 DEFINITION APO02553 BA000007
 ACCESSION APO02553.1 GI:13360211
 KEYWORDS
 SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
 DNA.

ORGANISM Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (sites)
 Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
 Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
 Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
 Sasaki,K., and Shinagawa,H.
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak
 Genes Genet. Syst. 74 (5), 227-239 (1999)
 2 (sites)
 Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
 Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
 Hayashi,T.
 Comparative analysis of the whole set of rRNA operons between an
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
 Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
 20557356

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 3 (sites)
 Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
 Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
 Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,K., and
 Shinagawa,H.
 Complete nucleotide sequence of the prophage VT1-Sakai carrying the
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
 O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)
 20564182

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 4 (sites)
 Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
 Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
 Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasaki,K.,
 Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M., and
 Shinagawa,H.
 Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)
 21156231

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 5 (bases 1 to 297816)
 Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
 Hayashi,T.
 Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
 Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
 Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
 URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
 Fax:81-6-6879-2047)
 genome project.

COMMENT
 FEATURES
 Source Location/Qualifiers
 1..297816
 /organism="Escherichia coli O157:H7"

gene
 CDS
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECs0753"
 79..1245
 /gene="ECs0753"

/note="similar to SUCC_ECOLI g111786948 percent identity
 100 in 368 aa (conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="succinyl-CoA synthetase beta subunit"
 /protein_id="BAB34176.1"
 /db_xref="GI:13360212"

/translation="MNLEHYQAKOLFARYGLPAPVYACTPPEAEASRIGAGPW
 VKQVHAGRGKAGGVKVVNSKEDIARAEMLGKRLVYOTDANGCPVNOILVEAT
 DIAKELYLGAADVDRSSRRVPMASSTEGVEIEKVAEETPHLHKVADPLGPMYOG
 RELAFKLGEGKLYOQFTKIFMGATITFERDLALIEINPLVITKQSGDITLDKGA
 DGNALPBDLEEMRDQSDPREDAQOMELYNALDNGICWNGAGLMTGTDIV
 KLHGEPANFELDVGGATKERTYEAKFTILSDKVKALVNIPEGIVKCDILAGGII
 AVAEVNPVYVYRLGNNALGAKKLADSGLNTIAAKGLTDAQOVVAVERK"
 1245..2114
 /gene="ECs0754"
 1245..2114
 /gene="ECs0754"

/note="similar to SUCC_ECOLI g111786949 percent identity
 100 in 289 aa (conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="succinyl-CoA synthetase alpha subunit"
 /protein_id="BAB34177.1"
 /db_xref="GI:13360213"

/translation="MSILIDKNTKVICGSGTSPHSBDATVYTKRMGCTPPGK
 GTFHGLIPFNVYREAAATGATSVIYVAPFKDSITLADIGLITITTEGIP
 LDMLTVYKLDDEAVRMIGPVGIVITPEGKDTIOPGHIHKPKGVIVSRGTLVE
 AVKQTDYDFGOSTCVIGIDPPTGNSNIDLEFEDDPQTEAIVMIGETGSAEEA
 LAATKEHVTKPVGVYIAGVYAPKGRMRMGAGATITAGCGTADEKFALEAAVKTVS
 LADIGALKIVLK"
 2781..3686
 /gene="ECs0755"
 2781..3686
 /gene="ECs0755"

/note="probable transcriptional regulator, similar to
 transcriptional regulators e.g. glycine cleavage system
 transcription activator (gcv operon activator) -
 Escherichia coli g11417043|sp1P32064|GCVA_ECOLI percent
 identity 31 in 300 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental

/product="putative transcription regulator"
 /protein_id="BAB34178.1"
 /db_xref="GI:13360214"

/translation="MRGKIPKTELVTFEVVYARHESYTRAARELATQSAVROVSAL
 EEPFRMPFSSHRSKRIFLNDTKYVIGYKETINKLEPRNTNTMTQPVQVTEIAVN
 LPFETILPVGSGSLAMSDOKLSVALELTELPLHGSTRTGTGEWFAISGVSSPLVN
 NGRPFDLSMLIAVNSNIGVALLPFFAIOHDDSDMWI PCDPVIRTRNTMTQW
 EKSDSPHLOFRMLAKSVPOEM"
 complement (3720..4322)
 /gene="ECs0756"
 3720..4322
 /gene="ECs0756"

/note="probable cob(I)alamin adenosyltransferase, similar
 to cob(I)alamin adenosyltransferases (cortinoid
 adenosyltransferases) e.g. [Escherichia coli]
 g1115148|sp1P13040|BRUR_ECOLI percent identity 67 in 200
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental

```

/product="putative cob(1)alamin adenosyltransferase"
/db_xref="GI:13360215"
/translation="MEARISTERYVOROKLEQVDYTVAAATEKGIIVTGNKNG
KSTAAAGTAVAGHGTGVAQYIKGMDNKEVNLDPVVERHIMGTGTTWTOR
QADIDAKESWSESKMLADRYDLVDELTYMLAHYLDTEVIAIIONRPAQOVS
IVRGCHSOILKMDYSEIRPKHAFDNGIOAPGIDW"
complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/translation="probable fumarate hydratase, similar to fumarate
hydratase e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g111205981sp|P00923|PUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPTWOPPELONKQDREYTLISDQHTVTELDGEYIKIAPE
ATLLARQAFYEASFLRSADHQQVASILNDQASNDKVALQILRNNEVAKGVP
NCODTGTATIVASKQDQWVGNDALSKGISTFOENLRFSQAPIDMTVEVNO
TNLPQDIDISAVAGDEYHFLCVNKGGSANKALYQETSLQPERLFLIEKKSL
GTACPPYHIAFVVGGLSADQTLAKLASTKYDNIPTSGNQGQAFDIDIEKVL
EASQCFIGAQFGGCFPAHDIVRLPRHSGSCPIAMALSCSDNRNINKHGLML
EKLHNGQYIPASIRENNAOHVOLDNRLRDVMDLARSLEPGTGVSGPIYVAR
DIAHAKIKARLDSCEPPEYIKHIIYVYAGPAPENMACGSGPTGGMDGYIDTF
QAGGSLVMSKSNRSDQVTDACHKGGFMUGSIGRAALAEYKSLRCLEPEPLG
MEAVMMMEVENLPALFVLDKGNFQFQDHKACSCPAHG"
complement(6092..7372)
/gene="ECS0758"
/complement(6092..7372)
/gene="ECS0758"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. (Bacillus stearothermophilus)
g11214671sp|P24943|GLT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMLALVLMIGVILNNTASPEYAKLYAOEISFT
TIFRLIKMTINPLVYSTLYGIAKMGDAKALGRFSTLFLFCASLSTALGLIV
NFPMPTGINTFAHGAETTVYAAPEPLKVFISAPPTSDVMAHNEILOIVFST
FLGCSITAIIGKGSALVHALDSLAHMLKLTGYWLPPLVPAISALAREGLAVM
VSAGIFMGEPFTMLLWLLGLAIVVGPILIRLIRLALSEPALATTSSSAAPP
GLTLEQGVSPKIASFVLPISGSEFNVGSMVCSFATVETACQCNHUSIGEDITM
LILMTSKMGAGVPRASMVVIAATLNOFINPEALILLMGVDFLDMGRSATVMSN
AMCAAMVSRWEEHGECCGKALKPNESNALP"
complement(7533..7850)
/gene="ECS0759"
/complement(7533..7850)
/gene="ECS0759"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. (Bacillus stearothermophilus)
g11214671sp|P24943|GLT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34182.1"

```

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcagcaggtgcatgcagcatt 25
|||||
Db 92389 GCGCAGCAGTGCATGCCAGCATT 92413

```

```

RESULT 14
E00893
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIROMO YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain-Escherichia coli Ns101;
CC Feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
FT /product="biotin synthetase".
FEATURES
source
location/qualifiers
1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN
Query Match 88.0%; Score 22; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 cagcaggtgcatgcagcatt 25
|||||
Db 123 CAGCAGTGCATGCCAGCATT 144
RESULT 15
MAV250020
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
MAV250020 2923 bp DNA BCT 26-SEP-2000
Mycobacterium avium subsp. paratuberculosis insertion sequence
IS900, Locus 7.
AJ250020
2599 gene: 2600 gene: insertion element; insertion sequence IS900;
p43 gene: transposase.
Mycobacterium avium subsp. paratuberculosis.
Mycobacterium avium subsp. paratuberculosis; Actinobacteridae;
Bacteria; Firmicutes; Actinobacteriales; Actinobacteriaceae;
Mycobacteriales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
1 (bases 1 to 2923)
Bull.T.J., Hermen-Raylor,J., Pavlik,I., El-Zaater,I.F. and Tizard,M.
Characterization of IS900 loci in Mycobacterium avium subsp.
paratuberculosis and development of multiplex PCR typing
Microbiology 146 (Pt 9), 2185-2197 (2000)
20431891

```

```

REFERENCE      2 (bases 1 to 2923)
AUTHORS       Bull.T.J.
TITLE         Direct Submission
JOURNAL       Submitted (29-NOV-1999) Bull T.J., Surgery, St.George's Hospital
              Medical School, London, SW17 0RE, UNITED KINGDOM
FEATURES      location/Qualifiers
              1..2923
              /organism="Mycobacterium avium subsp. paratuberculosis"
              /sub_species="paratuberculosis"
              /db_xref="taxon:1770"
              complement(11..466)
              /gene="2600"
              /complement(11..466)
              /function="unknown"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="CAB96064.1"
              /db_xref="GI:8919147"
              /translation="MYLVEIGTVDLPVKGAVNTILYFAYGAVLIVGYANDLIT
              PGRLQDLVIFDRPNAVVVAGMYATATVITTAIANSYSGGLGVAVYGLMGI
              LIGVALLAMHLLIPGSHFHEVEPOLHPGSAVALLLAVGVTAAVS"
              473..1922
              /organism="Mycobacterium avium subsp. paratuberculosis"
              /insertion_seq="IS900"
              /sub_species="paratuberculosis"
              /db_xref="taxon:1770"
              686..1906
              /gene="p43"
              686..1906
              /gene="p43"
              /codon_start=1
              /transl_table=11
              /product="transposase"
              /protein_id="CAB96065.1"
              /db_xref="GI:8919148"
              /translation="MTVEVYVAPVWAGVYAGACADHYCWITINDAQRLLSORVAND
              AALLLEIAVTTIADGGEVYVATIDNAGGALLIALLIAGORLLYIPGTVHHAAS
              VRECKTPAKDAIADQARMHDIPLRAGDDIVALELRTISRSULVDRPALEP
              NARPAAGILSALERAEDYKSRNALIILTYOTPDALRSAGARVAAFLLKRRARNAD
              IVAATLQANAHSHIVPGQDLATVVARLAKEVVALDTEIGDDAIEERFRRHRA
              EILSPGEVYLGAEFLAATGMAAFASADRLAGVAGIAPVROSGRTISGMLKRP
              RYDRLLRACYLSALVSIRTPSSRTTYDKRRTGKRHTOAVLALARRRLVLMMLR
              DHAYHPATTTAA"
              complement(11963..2394)
              /gene="2599"
              complement(1963..2394)
              /gene="2599"
              /function="unknown"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="CAB96066.1"
              /db_xref="GI:8919149"
              /translation="VSRNRLFEVIALAVAVVCLVTGIILOKNIASVYAGHYHYA
              RDVNGTRYQCTGSEVADTLEAEPAADNKTETRYSDIYIVPDNPTCSI
              RVPELSAGYSHRAVFEVLPGEFPGSGAGGTPGPGGK"
BASE COUNT    473 a 1005 c 1000 g 445 t
ORIGIN
Query Match   80.0%; Score 20; DB 3; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 cagcaggtgcacgcgcagca 23
    |||
Db 125 CAGCAGGTGCATGCCACGCA 144

RESULT 16
PSEHRPIA/c

```

```

LOCUS      PSEHRPIA      3603 bp      DNA      BCT      21-OCT-1993
DEFINITION Pseudomonas syringae hrp1 genes, complete cds.
ACCESSION  U11582
VERSION    U11582.1
KEYWORDS   GI:151267
SOURCE     hrp1 gene; protein translocation.
ORGANISM   Pseudomonas syringae (strain 61) DNA.
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE  1 (bases 1 to 3603)
AUTHORS   Huang,H.-C., Xiao,Y., Lin,R.-H., Lu,Y., Hutcheson,S.W. and
            Collmer,A.
TITLE     Characterization of the Pseudomonas syringae pv. syringae 61 hrp1
            and hrp1 genes: homology of Hrp1 to a superfamily of proteins
            associated with protein translocation
JOURNAL    Mol. Plant Microb Interact. 6, 515-520 (1993)
FEATURES   location/Qualifiers
            1..3603
            /organism="Pseudomonas syringae"
            /db_xref="taxon:317"
            1330..3417
            /gene="hrp1"
            1330..3417
            /gene="hrp1"
            /note="hypersensitive response and pathogenicity gene,
            membrane-spanning protein; hrp1 superfamily proteins are
            associated with protein translocation; putative"
            /codon_start=1
            /transl_table=11
            /product="Hrp1"
            /protein_id="AAA03015.1"
            /db_xref="GI:151268"
            /translation="MNRVINFNLNVALSAMRSELVGFVAVIAIVEMITPLPTGLID
            VLIANICISLIMLHMLPRPLAFSPFPAVILLTMEPLASVSTRILLNONG
            HIVEAFGEVVGNLAVGLIVFLITVYNLVITKGERAEVGAFTIDAMPKQMS
            IDSDDRANLITVHARRKRAETKESOLFQMGAMKFGVGDIAISLIIVAINMIGI
            SIGVQHNMAAGDQLQYTVITIGDGLIADIPALLISVSGMITVVPNTAEVANI
            DEQRTAEQTSPPKMITASVAMLEFALPMPGVETTAIIICGAGLQLOAKAKRA
            GUTLPSEITIEHVDIADDERFVYDVPMKATFTQSHVAVEAROLEGNLPAIIPGN
            TDREDDQWMLPABOSGELNPNVSTTLIRMRALOSCAPOTIGIOETKATISMES
            BOPELAGQMGVLTITRESAVLORLAECVPLAIVIAETLITRHOQERDINLVTDY
            VRIALSOIYHOYCGAGIOWMLTTPSESLIDGIRQOTETFPALSWETSOMLVQO
            LHAIPVAPQOAVLVAODLRSLRLTLKEETTHVPLSFATISNAKVKWGRDIL
            EDLLEPLDNEHAA"
BASE COUNT    772 a 1093 c 1035 g 703 t
ORIGIN
Query Match   76.0%; Score 19; DB 3; Length 3603;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 gcaggtgcacgcgcagcat 24
    |||
Db 1522 GCAGGTGCATGCCACGCA 1504

RESULT 17
MTCY180/c
LOCUS      MTCY180      44201 bp      DNA      BCT      17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
ACCESSION  Z97193
VERSION    Z97193.1
KEYWORDS   GI:3261816
SOURCE     Mycobacterium tuberculosis.
ORGANISM   Mycobacterium tuberculosis.
            Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
            Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
            1 (bases 1 to 44201)
            Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eigemeier,K., Gas,S., Barry III,C.E.,

```

TITLE
 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
 JOURNAL
 Nature 393 (6685), 537-544 (1998)
 MEDLINE
 98295987
 REFERENCE
 Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]
 2 (bases 1 to 44201)
 AUTHORS
 Parkhill, J.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2225942.

COMMENT

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in Tbpase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (arg, gtg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source
 1..44201
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 <1..42142
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 /clone="Y180"
 23..460
 /gene="Rv1873"
 23..460
 /gene="Rv1873"
 /note="Rv1873, (MTCY180.45c), len: 145. Unknown, Tbpase score is 0.909"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1873"
 /protein_id="CA10052.1"
 /db_xref="GI:2225987"
 /db_xref="SPTREMBL:O07756"
 /translation="MKSASDPDLKRFVYAAQAPVRSVEELIRAGKRGHMMFVFPD LKLGSSPLFARVYGISLEAOAYLOHDLIGRLHECTGLVNOVGRISEIFGPPD LKICSSMTLFAARTDANQDFVALLAAYYGGSDRRFVALAVT"
 517..520
 /note="possible RBS, GGAG, for Rv1874"
 533..1219
 /gene="Rv1874"
 533..1219
 /gene="Rv1874"
 /note="Rv1874, (MTCY180.44c), len: 228. Unknown, Tbpase score is 0.928"
 /codon_start=1
 /transl_table=11

/product="hypothetical protein Rv1874"
 /protein_id="CA10067.1"
 /db_xref="GI:2225986"
 /db_xref="SPTREMBL:O07755"
 /translation="MLMRPDDDCAGQARQAVADALLGIGVAGLSTINVRDSTL MTLTLPVPAVAVSLMTQOCGEQVAAALRLAOBCDELATYTESVPLTFPSIVE GSGRPTGLANIALIRPDLDOATWLTWRQDHQVATLEAOATGTYQNRVRLATE APAGTIVEELFPVAVATDLKAFGAADDNDLRNISRMAVSTAFGANONIDIVPTS RYKRTPPFD"
 1218..1221
 /note="possible RBS, GAGG, for Rv1875"
 1230..1673
 /gene="Rv1875"
 1230..1673
 /gene="Rv1875"
 /note="Rv1875, (MTCY180.43c), len: 147. Some similarity to G95384|MTC165.22 (147 aa) opt: 178 z-score: 240.3 E(): 7.4e-06; 26.98 identity in 130 aa overlap. Tbpase score is 0.908"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1875"
 /protein_id="CA10051.1"
 /db_xref="GI:2225985"
 /db_xref="SPTREMBL:O07754"
 /translation="MTTLNEAALAAEAERGLAVSTVRADGTVOASLVNGLPVPV GEPGLGTYTGKVLGNLRAFPOLATFERNQWMAVEARQALVGPDPRLVDGER LRLIRVFVAGAGTHDDMDXDYRVMAQEQRAVAVLTPTIYSNG"
 2179..2183
 /note="possible RBS, AGGAG, for Rv1876"
 2189..2668
 /gene="bifA"
 2189..2668
 /gene="bifA"
 /note="Rv1876, (MTCY180.42c), bifA, len: 159. Function: bacterioferritin (bfr) (159 aa) opt: 958; E(): 0; 90.68 identity in 159 aa overlap. Tbpase score is 0.913"
 /codon_start=1
 /transl_table=11
 /product="bifA"
 /protein_id="CA10050.1"
 /db_xref="GI:2225984"
 /db_xref="SPTREMBL:O08465"
 /translation="WQSDPDLRLNEQLSELTAIINOYFLHSMQDNQGETLAHT RAESDEMRHAEITDRILLDLGPNYQISGLNIGTLLRQEPADALAEYVDNLK PGIVMCREKQDTSVAVLEKIVADEEHIDYLETQLEMLKGLSELYSACVSRPPI"
 2742..2746
 /note="possible RBS, GGAGG, for Rv1877"
 2753..4816
 /gene="Rv1877"
 2753..4816
 /gene="Rv1877"
 /note="Rv1877, (MTCY180.41c), len: 687. Integral membrane protein similar to many antibiotic and drug efflux proteins. M. tuberculosis relatives include: MTCY3012.01, MTCY98.0002c and MTCY369.27c. Contains P500217 Sugar transport proteins signature 2 (P500217). FASTA results: 056175 (357 aa) opt: 895; E(): 0; 34.78 identity in 528 aa overlap. Tbpase score is 0.916"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1877"
 /protein_id="CA10049.1"
 /db_xref="GI:2225983"
 /db_xref="SPTREMBL:O07753"
 /translation="MAGPTAPPTAPPAIRAGPLSPYRNITITATPAGVIAATGQ TIIVPALPTIVAEIGSTVDQSWAVSYLGGTVVAVVAGKGLDLGRNRVILGSAVF VVSGVLCGLSOTMTLAIISRALQVAGAGIYSTAVLAIAVPLDRKRYGAGAVF GNVNTPGILGLTIDYLSMRNAFNINPVSIJAVLVATAVPAALARPKEVLDYGLI LVIAVATTAIILATISWCGTVMGSAFTIYGLILGAIVAGFVMEGRRAAAILPRL FGSFPAVCCVLSFVYGFANLGLTFPVPYICGYVGSATGASLTPLVIGLILAST GTGLVAGRTGRYKIFPVAGMALNAVAFILMSQMDERTPLLSLTLVVLGAGIGLSMO

VLVLIVONTSSFEEDGVATSGVTFEVRVVGASFGTATFGALFVNFELDRIGASLTGSGAV
 PVPAPSPAVLHOLPQSMAPITVRAVESLTOVLCASVTAVGFIALLREVPLTD
 IHDADDDLDGDEGVRAESPEDVLEIAVRRLPBGVRLIATPOCGGAVATALL
 RIYOURLFEAVRRLDGRHLHPVGVFEVPEPDLNOTGTGAARDLITLTGSHKOV
 DSIATVLIKRWLIDHLAVAPGLKRPDPHGFALAOHTVDAVLDVORDWEDLDGLESRO
 LAATP"

misc_feature
 3122..3199
 /gene="RV1877"
 /note="PS00217 Sugar transport proteins signature 2"

gene
 4871..6223
 /gene="glnA3"
 4871..6223
 /gene="glnA3"
 /note="glnA3"
 /note="RV1878, (MTCY180.40c), len: 450, glnA3, similar to
 many glutamine synthetases. FASTA results: GLNA_BACCE
 P19064 (EC 6.3.1.2) (443 aa) Opt: 497; E(): 5.2e-23; 29.0%
 identity in 331 aa overlap. Tblast score is 0.955. Also
 similar to C-terminus of FLUG_EMENT P38094 flug protein.
 emericellulidians (865 aa), fasta scores: opt: 227
 z-score: 357.9 E(): 6.4e-13 (29.9% identity in 394 aa
 overlap) (note that the downstream ORF MTCY180.39c is
 similar to the N-terminus) and to several other M.
 tuberculosis proteins eg. MTCY427.03c, MTCY427.01 and
 MTCY190.31"
 /codon_start=1
 /transl_table=11
 /product="glnA3"
 /protein_id="CAR10048.1"
 /db_xref="GI:2225982"
 /db_xref="SPTREMBL:O07752"
 /translation="MTATPLAAAIATOLEAEAGVDVIVVPAQLQAKTPIRRIN
 TEANPLGASPVWTFECIDCSIAFTADISVGDRLRIDLSARITGDIAMPACF
 FEOGTPVPACSRGTLSRIEALADAGIDAVIGHVEFLIDADQCRPSTLMQYV
 AGVLEHFAFVDVNAATAGATATGEPHPCANFEISLAPVPAADLVLTLL
 IGSTARHGLRVLSPAPRASIGSGAHOHRSULTMSEGLFSGGTGAAGMTSAEAAV
 AGVIRGLPAGCIIIGSIVSGIRMRPGMAGITACVSETEVPTOLSDVDRACITRL
 NYEVKVVDSANPPLASAVILGLADGKTKTAIVPESETVPTOLSDVDRACITRL
 ADQADATAVIDSKILRCIIDDVVAIVAVVROLEHERYGLDPLPAALADKFRNANV
 6213..6216
 /gene="glnA3"

RBS

Query Match
 Best Local Similarity 76.0%; Score 19; DB 3; Length 44201;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgcacgcga 20
 |||||
 Db 25769 CGCAGCAGGTGCATGCCA 25751

RESULT 18
 LOCUS AC068068/c 37500 bp DNA HTG 28-APR-2000
 DEFINITION Leishmania major chromosome 2 clone L5666 strain Friedlin, ***
 ACCESSION AC068068
 VERSION AC068068.1 GI:7658330
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Leishmania major.
 ORGANISM Leishmania.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 37500)
 MYLER,P.J., SISK,E., Ruiz,J., Cosenza,P., Cruz,A. and Stuart,K.
 Direct Submission
 Submitted (28-APR-2000) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 12249: contig of 12249 bp in length
 * 12250 12296: gap of unknown length
 * 12297 27295: contig of 14999 bp in length
 * 27296 27342: gap of unknown length
 * 27343 27885: contig of 543 bp in length
 * 27886 27932: gap of unknown length
 * 27933 28468: contig of 536 bp in length
 * 28469 28515: gap of unknown length
 * 28516 28657: contig of 142 bp in length
 * 28658 28703: gap of unknown length
 * 28704 29418: gap of 715 bp in length
 * 29419 29464: gap of unknown length
 * 29465 30625: contig of 1161 bp in length
 * 30626 30671: gap of unknown length
 * 30672 31824: contig of 1153 bp in length
 * 31825 31870: gap of unknown length
 * 31871 33509: contig of 1639 bp in length
 * 33510 33555: gap of unknown length
 * 33556 35178: contig of 1623 bp in length
 * 35179 35224: gap of unknown length
 * 35225 37500: contig of 2276 bp in length.
 location/Qualifiers
 1..37500
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="L5666"
 /chromosome="2"

BASE COUNT
 ORIGIN
 6965 a 11364 c 11233 g 7435 t 503 others

Query Match
 Best Local Similarity 72.0%; Score 18; DB 73; Length 37500;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgcacgc 18
 |||||
 Db 5534 CGCAGCAGGTGCATGCC 5517

RESULT 19
 LOCUS AC010079/c 38000 bp DNA HTG 11-SEP-1999
 DEFINITION Leishmania major chromosome 2 clone L4116 strain Friedlin, ***
 ACCESSION AC010079
 VERSION AC010079.1 GI:5870277
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Leishmania major.
 ORGANISM Leishmania.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 38000)
 MYLER,P.J., SISK,E., Ruiz,J., Cosenza,P., Cruz,A. and Stuart,K.
 Direct Submission
 Submitted (11-SEP-1999) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 1078: contig of 1078 bp in length
 * 1079 1201: gap of unknown length
 * 1202 22879: contig of 21678 bp in length

FEATURES * 2280 23001: gap of unknown length
 * 23002 38000: contig of 14999 bp in length.
 Location/Qualifiers
 source 1. 38000
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /chromosome="2"
 /clone="L4116"

BASE COUNT 6861 a 11830 c 11904 g 7132 t 273 others

ORIGIN

Query Match 72.0%; Score 18; DB 61; Length 38000;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgcatcgc 18
 |||
 Db 16164 GCGACAGCGTGCATCGC 16147

RESULT 20
 AC087836 42000 bp DNA HTG 28-FEB-2001
 LOCUS Leishmania major chromosome 2 clone B4162 strain Friedlin, ***
 DEFINITION SQUENCING IN PROGRESS ***, 4 unordered pieces.
 AC087836
 VERSION AC087836.2 GI:13162520
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Leishmania major.
 ORGANISM Leishmania major
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 42000)
 REFERENCE 1
 AUTHORS Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K.,
 Dodson,D.E., Beverley,S.M. and Scholtes,L.
 TITLE Direct Submission
 JOURNAL Submitted (30-JAN-2001) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 On Feb 28, 2001 this sequence version replaced gi:12597743.
 COMMENT * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 8057 8056: contig of 8056 bp in length
 * 8290 8289: gap of unknown length
 * 18959 18958: contig of 10669 bp in length
 * 19191 19191: gap of unknown length
 * 40819 40819: contig of 21628 bp in length
 * 40820 41051: gap of unknown length
 * 41052 42000: contig of 949 bp in length.
 Location/Qualifiers
 source 1. 42000
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /chromosome="2"
 /clone="B4162"

BASE COUNT 7450 a 12790 c 13550 g 7510 t 700 others

ORIGIN

Query Match 72.0%; Score 18; DB 77; Length 42000;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgcatcgc 18
 |||

Db 34508 GCGACAGCGTGCATCGC 34491

RESULT 21
 AE004192
 LOCUS AE004192
 DEFINITION Vibrio cholerae chromosome I, section 100 of 251 of the complete
 chromosome.
 AE004192 AE003852
 AE004192.1 GI:9655581
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Vibrio cholerae.
 Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

REFERENCE 1 (bases 1 to 12891)
 AUTHORS Heidelberg,J.F., Eissen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
 Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
 Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
 Ermolaeva,M.D., Vamathevan,J., Bacs,S., Qin,H., Dragoi,I.,
 Sellers,P., McDonald,L., Utterback,T., Fleischmann,R.D.,
 Niemann,W.C. and White,O.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 LOCATION/Qualifiers
 1. 12891
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="01"
 /db_xref="taxon:666"
 /chromosome="I"
 /note="biotype: El Tor"
 /complement(84..1370)
 /gene="VC1111"
 complement(84..1370)
 /gene="VC1111"
 /note="similar to GB:J04423 SP:P12995 PTD:457106
 PTD:490217 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=1
 /product="adenosylmethionine-8-amino-7-oxononanoate
 aminotransferase"
 /protein_id="AAE94270.1"
 /db_xref="GI:9655582"
 /translation="MMMDLAFDRCHIMHPYTSLLPRLCYPVSRADGWTLEDRSL
 VDGSMMMAALHGYNHHPKLNQAAHQIEQMSHIMFGSLHNPALTELCOQLKLAIPNNL
 QOVLADSGSIVAYEVSLKMALOYHSGQPRAFELRLHNGYHDDPFAASVDDPNSM
 HSLYKGLPRLHIFANSPEGGFPADMDRODADRNLHTENNQDAIVLEPIYOGAGG
 MRYIHPERLQVRLACDEGVEVLLIDELATNGRRPKLRACSEAGIOPRIICGVKALT
 GGWYTLSATLTQQVADPTVCAEGAGCGCMHGPRTMGKRLACAYASATSLIEGDMDQO
 VANIETFAEDELPRINDSPRYKQTRWGAIGVETIDLPVNMVEYQLAEVHGWIIRFP
 GRLLYIMPPYISQPEHLQKLVDMMAALNOPDFKQ"

gene
 CDS
 1497..2549
 /gene="VC1111"
 1497..2549
 /gene="VC1111"
 /note="similar to GB:J04423 SP:P12996 PTD:145425
 PTD:490219 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1


```

/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEYRHNWVAEVKALLDKPEMILLFEAQOVHRLHPPHNOVST
LSTKAGCEPDCYCPQSAHYRTDVKEERLMEVERVLDAQAKNSGSTRPCGAAM
KNPERMDPLKEMI RGVKDMGLETCMTLGMTPDOAOQLAAGLDYINNLDPSPF
YGNITTRTODRLDITLSHVADGAK ICSCGIIOMGESTNDRAGLVEIANLPHPES
VPIMLVKVGTPLQGVDDYEPFVRLIVARIMPKSAVRLSAGREKNEQOALC
FMAGANSIEFGCKLITTPNPAPEDSDMLFFKRLGINRQVAKFDEITENLDRVER
VAARPTASDLFQDAL"
2533.3687
/gene="VC1113"
2533.3687
/note="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRKRSRIEDALNREHSRLTRQLLDRSAQGEILQODRTYL
MESSNDYIGLANAEILVQAMQGLARFGAGSGSPWVTGSSNAHAELEHCHWIGYE
RAVLSSGSSANQALLFTLEKEDLLIDKLNHASLMEAGMLSPATMKRPFKINDPEHL
ROLLEHOSNLSLVTEGVFSMDGCAPLQAIATLVKORAWLMDADAGTGVIGEDAGS
CQAAGIHPELIVYTFKAFAGLGAVALCAHAGDYLTORARHHYVSTAMPQOARHLT
HAARMIOSMSREQLAEELACEDDCCRNIPGFVATOTPIKMWLGSSSALSASHEL
KOOGIWSAIRPTVPVGSARLRTITTAHTOQIRLSEQLHWEKDT"
3549.4487
/gene="VC1114"
3549.4487
/note="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthesis protein B10c"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MGECSTHRCROCATFYHNGCATYPSADTPLLGAVKTCDEEG
HMSMTATELCEVLKKSALAOAFKAAHNDHAFQDVGLRLQKMSCLKGLRVL
DGGCTGYRSALIRERGAQVCCADISHMLDQAKORCGDEGMYOLADAEOLPFASAC
EDWFSLSLQWCEDELSLPLSEIRVLPHGQAFSTLSDSLELEQAMRSVDHHRH
INQITISNOVKIALAQAGCSQHIDLAATYWEAFAWLMMDKIGCANHYSGSTGL
ISRRLAKVEALVQSFRRNOGLVPATYVQVGLIHR"
4463.5179
/gene="VC1115"
4463.5179
/note="similar to GB:S68059 SP:P13000 PID:1490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="gethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSPPRNAFIAGTDTVDGKTVASKATLQALQALHNTATTY
KPVASGDKTEFGYRNSDALHMAKATVDYVNVYALVLPSPITAKHKEVTTD
YALTSNLSKRONAEILVYVAGAGWVPSTSDCSLTSWKQERLPIVLAVKLGGL
SHALTAEAIRADGELVCMANRINPTEHTATIEHLBGRGSTPLGELPYPMKAK
ROELGKFIQDLHLEPPTVA"
complement(5261..5470)
/gene="VC1116"
complement(5261..5470)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

```

/gene
CDS
/transl_table=11
/product="heat shock protein HspX"
/protein_id="AAF94276.1"
/db_xref="GI:9655588"
/translation="MSRNLTTNKGKHNMAEASHYQMAIDLLCCHLGISEQEARQDGL
EAKETPLQOOIYDTHSLMGLSAEK"
5697.6560
/gene="VC1117"
5697.6560
/note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="heat shock protein HspX"
/protein_id="AAF94276.1"
/db_xref="GI:9655588"
/translation="MKRIILFLATNLAVVLVSVLNTVAVTGMQSGSLGLVMAA
VFGFGAFSTILMSKSMALRSYGVVDTTPRNEHEMLLETVRQANDAGIGMPTVAI
YDAPDMAFATGAKRDSLVAVSTGLNNMTDEAEVLAHEVSHLANGVMTLMQ
GVNTFVFLSRFTANIVASRDSERGSNNMYEVSVMVLGTFEGLASFITMYSR
HREFHADGAQALVGKHKMIALERLKKQGESHLEGSMMAPGIGTKRSLSELMVTHPP
LEKRLALRNM"
6770.7198
/gene="VC1118"
6770.7198
/note="similar to PID:1653608; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AAF94277.1"
/db_xref="GI:9655589"
/translation="MSIMDVQVNAQFOQLDKSQLRLIEIHPDVFEEDAAHRIEGF
DALQYFLNLYQNVHTCTFTIHQYVANECAFVMTNHLHPKLAGEQVDVAGVSHL
HPAGKVTYHRDYEDMGEMYLEQPLVIGVYIRAKRRLQO"
7195.7920
/gene="VC1119"
7195.7920
/note="similar to SP:P25970; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AAF94278.1"
/db_xref="GI:9655590"
/translation="MTAVFTTGATSGIGKOLADYAKQSGNOYIACRNSVLSLHTO
YANIFPLAFDVTDPGTKALAOPLCOPELMTLNAGDCXYIDGKMDYTLMARVRIN
VLGVAAVTEGLOPHLSGHRVATVGSIASELAIPRAEYAGSKAAVAYIARTLQDMR
PLGIEVTTIFPGFVATPINDRNTFAMPPIITTVRAQETIKAGIARGSVSLYPPKRPW
LIRLLGAPRAMQGRIVRLRKA"
7931.9211
/gene="VC1120"
7931.9211
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

```

Query Match 68.0%; Score 17; DB 1; Length 12891;
 best local similarity 100.0%; Pred. No. 29;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgacgagtgatcgatc 17
 Db 1575 GCCGACGAGTGATCG 1591

RESULT 22
 AK024397 4674 bp mRNA PRI 29-SEP-2000
 LOCUS AK024397/c
 DEFINITION Homo sapiens cDNA FLJ14335 fis, clone PLACE4000411, highly similar

to Homo sapiens mRNA: cDNA DKFp586D0624 (from clone DKFp586D0624).

ACCESSION AK024397
VERSION AK024397.1 GI:10436775
KEYWORDS oligo cloning; fls (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_1lb:PLAC4 clone:PLAC400411.

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagaetsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T. NEDO human cDNA sequencing project

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 4674)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Katsuragi, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert International Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES
source
1..4674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLAC400411"
/clone_1lb="PLAC4"
/issue_type="placenta"
/note="cloning vector: pME18SPL3"
608..1273
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14908.1"
/db_xref="GI:10436776"
/translation="MLFSKYLCMTVEOPKGEELSAAIKRIYATKASGKKLQKVT LKSPRGLITDNLITNLIENSVISYCTADKMDKVAAYIAOSHOHSLCHAF L CYRKMAQAVTLVAOAFKVAPEFMVSKSEKRDKASQEGDVGARQDCTPP LKS LVATGNLDLEETAKAPLSTVSNATNTMDEVPKPPALSGSSVYVSGVASRSVILSL TSG"

BASE COUNT 991 a 1342 c 1239 g 1102 t
ORIGIN

Query Match 64.08; Score 16; DB 89; Length 4674;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 ggtcgcgcgcagcat 24
|||||
Db 2446 ggtgcgcgcgcagcat 2431

RESULT 23
AE005685 10290 bp DNA BCT 28-MAR-2001
LOCUS AE005685
DEFINITION Caulobacter crescentus section 11 of 359 of the complete genome.
ACCESSION AE005685.1 GI:13421206
VERSION AE005685.1
KEYWORDS Caulobacter crescentus.

ORGANISM Caulobacter crescentus
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.

REFERENCE 1 (bases 1 to 10290)
AUTHORS Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Kouri,H., Shetty,J., Berry,K., Uitterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)

TITLE Complete genome sequence of Caulobacter crescentus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
MEDLINE 21173698
REFERENCE 2 (bases 1 to 10290)
AUTHORS Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Kouri,H., Shetty,J., Berry,K., Uitterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C. and Fraser,C.M.
Direct Submission
JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1..10290
/organism="Caulobacter crescentus"
/db_xref="taxon:69394"
complement(74..1363)
/gene="CC0102"
complement(74..1363)
/gene="CC0102"
/note="identified by match to TIGR protein family HM TIGR00563"
/codon_start=1
/transl_table=11
/product="RNA methyltransferase Rmb"
/protein_id="AAK22089.1"
/db_xref="GI:13421207"
/translation="MTOELNDGLPAREALTLIDAAISRRGGLEDAASANGFRELPEP EAPFARALAMAVLRHLGIDRLATRLTOKAPDPVYMLLRGAQAFLEVPAPAAVA QVAGLANKTSRPRKGLVNAVLRGLDAPADPPSLRPPMYARVAVAFGEATP QIAGOTATPATDLSAKGADLAPLYEALAEVLPGLRLRRKGVAAVAMPGEFTW WQDDAAAVPARLDVQPGASVLDLCAAPGGKTLDLAAGAOVVALDLSAARLRISL NLRGLSKEIVADAAVWDDTRFDALILAPCSATGTFRRHPVILMAARGDVSL NAVQRIILDAADRLKPGARLVYCVSLEPEGSEAOVAFARODVLEPITAAEAP SAGPESATLTPRGTLRLPHRRDDGQDFFAARFKL"
1426..1704
/gene="CC0103"
1426..1704
/gene="CC0103"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22090.1"
/db_xref="GI:13421208"
/translation="WIVGGODRSISTRGAAMTDTPPEPDAEVEGAAPGKTLTP AARALFEAAARAALSLPREGGRGDPPTFRGDMKKGVAVDF"
1738..3489
/gene="CC0104"
1738..3489
/gene="CC0104"
/note="identified by match to TIGR protein family HM TIGR00066"
/transl_table=11
/product="gamma-glutamyltransferase"
/protein_id="AAK22091.1"
/db_xref="GI:13421209"

[illegible]

CDS /gene="CC0109"
 complement(6796. .7473)
 /gene="CC0109"
 /note="identified by match to PFAM protein family HMM
 PF02190"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent protease Ia domain protein"
 /protein_id="AAK22097.1"
 /db_xref="GI:13421214"
 /translation="MPVARKIDDLPLVLPVPLDGVLLPSGOLPNIFFERYPLNMI
 DDAMSGRMIGMIGTRPLPGGKGPORALAPVGCAGVTSFAETSDGRYLITLTLGRL
 RRTGGELVPRTPYRQVADAPYQADLRBAAGRTSETLRLMTALRRVLDHRLGL
 IDMSAEAPSDALINSLAMALPDPMEKQALLERATLFEKKATLTALLEIDAAGDDI
 DEPSIQ"
 complement(7489. .8388)
 /gene="CC0110"
 complement(7489. .8388)
 /gene="CC0110"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="Chlorodioxin"
 /protein_id="AAK22097.1"
 /db_xref="GI:13421215"
 /translation="MSLIGKEPAPSSKSDHIKGDGDSFMDVIEASKTOPVIVDFMFA
 TWGCPQRLPTPTIEKVYPAKGAQKIVYIDVKGPGAGLRSISPIVYAFVYGRPRP
 DAPGALDESQVRAFYVEKLACAGCALDEITANCKESLETGDIAGAAQAFQALQID
 PTKSAIGGAKRVYLVANGDEPGAEEVYAMAPADKDPDLARALALAAAFSETAA
 FEGRVADPDDEHARLELAKLAGVGRLDAAHLTLTIARDRTWMDAARKOLLTFVE
 EAAGPTSEVAKGRRKLSILFS"
 complement(8498. .9004)
 /gene="CC0111"
 complement(8498. .9004)
 /gene="CC0111"
 /note="similar to GP:6899999; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="DNA-binding protein, putative"
 /protein_id="AAK22098.1"
 /db_xref="GI:13421216"
 /translation="MKTTRADLFAFEDAGVDKTLIDHPYFVREEGLEITAAFGHPT
 KNLGLDAKQGLWLISALGETTIDLKRIHVTGSGRISFGQEMMLETLVTPSSVTA
 EGLINDTEKRRFVLDKALADSDPVNPHLKNATMTAVSOGLRRFLAALGVPEMIVD
 FAAEAYVG"
 9081. .9827
 /gene="CC0112"
 /gene="CC0112"
 gene
 Query Match 64.0%; Score 16; DB 1; Length 10290;
 Best Local Similarity 100.0%; Pred. No. 1e-02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1 ggcgcagcagcgcgc 16
 |||||||
 Db 1112 GCGCAGCAGGTCATC 1127
 RESULT 24
 MTW031 13350 bp DNA BCM 17-JUN-1998
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 6/162.
 ACCESSION AL021926 AL123456
 VERSION AL021926.1 GI:3261520
 KEYWORDS
 ORGANISM
 Mycobacterium tuberculosis.
 Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 Actinomycetales; Corynebacterinae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 13350)
 Cole, S.T., Brosch, R., Parkhill, J., Garfield, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,


```

/gene="RV3798" (MTV026.03), len: 444. Putative
/Note="RV3798, (MTV026.03), len: 444. Putative
transposase identical to ISI557 MTCY373_32,
sp1010621.Y02X.MTCU HYPOTHEICAL 50.2 KD, PROTEIN CY37
(444 aa), 99.8% identity in 444aa overlap and similar to
many, e.g. gp10576491.DB057649_1 Dibenzoifuran-degrading
bacterium (163 aa), FASTA scores: opt: 767 z-score: 926.8
E(): 0, 67.3% identity in 168 aa overlap, or to BRIS1001.1
B.parapertussis insertion sequenc (406 aa), 24.9% identity
in 402 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3798"
/protein_id="CA17862.1"
/db_xref="GI:2950417"
/db_xref="SWISS-PROT:O10621"
/translation="MKNVRLPALLGVKRTVIDIEEEDDAGGARVAVRPSA
VLRGCRGRKASWIDRAGLRKMSLDMGWVEVFLERARVNCPTGPTVAVPVA
RHHAGRTAFDDTVAMAVACSTAVCELMRYAMRTGATVARYAWATDEKIDIFANL
RRIGIDELSYKRRHRYLTIVVDHDSGLVMAAPGHDKATLGFDAAGAEKRAOITIV
SADADWIMADVETRCPDALOCADPEFVAVANDEALDERRRANNDARLARTPKGS
RGPGRVKNAPRGREARLRKARVALMKNPEDLTEROSAKLAMIADTPELTXAYLL
KESIRHVSNGKEEGKALDRWISMAORCRIPVPELALARKRHRVALDALDGLSQ
GLESNTIKRILTRIRAFGRFPALILAMLTLAGHRTPLGRHNPISQ"
1980..3314
/gene="RV3798"
/complement(3382..4935)
/gene="accda"
/complement(<3382..4935)
/gene="accda"
/Note="RV3798c, (MTV026.04c), accda, len: 517. Probable
propionyl-CoA carboxylase. Beta chain, similar to many,
e.g. sp1530031.PCCB_SACER PROPIONYL-CoA CARBOXYLASE BETA
CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E():
0: 50.7% identity in 519 aa overlap, Also similar to
u00012.17 M. leprae cosmid B1308 (549 aa), 49.8% identity
in 514 aa overlap and to MTCY71.20 M. tuberculosis cosmid
(548 aa), 48.9% identity in 515 aa overlap"
/codon_start=1
/transl_table=11
/product="accda"
/protein_id="CA17863.1"
/db_xref="GI:2950418"
/db_xref="SPTREMBL:O53578"
/db_xref="SPTREMBL:O53578"
/translation="MTVTEVHTHTTAEKLAERLELAKKEGGKAAKRRKKIPS
ARARIELVDPGSFMEIGALCRTPDPNMLYSGVVTGGLNGRVCVFSDQVGS
GTVGMGKRVKAVLEMCAWGCPIVIGVDSGARIODAVTISLAMYALEGRHLLSG
LVPOISILIGKAGAVYSPITDVLVAVADQGYMFTVDVTKVGTGDEYSLDELG
ADHPASGNTHOVSEEAAYOVYRDLFLSPNCDPKPPVNPGLPEPTSHLELD
SIYDSDNMAVDMHEVILRTFDDGDFLDIAAQAQGLITGVARVGRITVAVANCPM
MSGATIDNASDKARFTRSDAFLPLVAVDTPGLPVEDEKNGITKRGRTFLYAV
VEADVPRVTTIRKSYGAVAVAGSKOLADLFAFPTARLAVTGADGAOULMKRFP
DPNAPENQAIRKSFVENYNLMAIPWIAERGFIDAVIDPHETRLLRKSMHLRDKO
LMWVEQRRKHG"
/complement(4932..10133)
/gene="pks13"
/complement(4932..10133)
/gene="pks13"
/Note="RV3800c, (MTV026.05c), len: 1733. Probable
polyketide synthase, similar to many, e.g.
gp1M63761SERERYA.1 S. erythraea first ORF of eryA gene,
involved in complex polyketide formation in erythromycin
biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E():
0, 33.6% identity in 1232 aa overlap or to AF040570.9
Amycolatopsis mediterranei rifamycin biosynthetic gene
cluster (3413 aa), 34.3% identity in 1162 aa overlap.
Also highly similar to mycobacterial putative polyketide
synthase: MTCY338.19, U00023.3 M. leprae,
MTCY22610.2, MTCY19H9.2, U00023.6 M. leprae, contains
P500606 Beta-ketoacyl synthases active site"
/codon_start=1
/transl_table=11
/product="pks13"

```

```

Query Match          64.0%: Score 16; DB 3; Length 23740;
Best Local Similarity 100.0%: Pred. No. 86;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2  CGCAGCAGGTCATCG 17
Db      3422 CGCAGCAGGTCATCG 3437

RESULT 26
LOCUS   POL245436/c
DEFINITION
Pseudomonas putida OCT plasmid alk genes cluster (alkBGHJLK, alkN
and alkST genes) and flanking DNA, strain Tfr-1L (Gpol), formerly
annotated as Pseudomonas oleovorans.
ACCESSION
VERSION AJ245436.1 GI:5824135
KEYWORDS
acetyl-CoA synthetase; alcohol dehydrogenase; aldehyde dehydrogenase;
alkane-1 monooxygenase; alkB gene; alkBFGHJKL regulator; alk gene;
alkS gene; alkT gene; hypothetical protein; methyl-accepting
chemotaxis protein; ORF; outer membrane protein; plasmid; recA
gene; recA gene product; rubredoxin 1; rubredoxin 2; rubredoxin
reductase; trpA gene; transposase.
SOURCE
Pseudomonas putida.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 34902)
Kok, M., Oldenhuis, R., van der Linden, M.P., Raaij, P., Kingma, J.,
van Ielyveld, P.H. and Wiltolt, B.
The Pseudomonas oleovorans alkane hydroxylase gene. Sequence and
expression
J. Biol. Chem. 264 (10), 5435-5441 (1989)
MEDLINE
89174501
REFERENCE
2 (bases 1 to 34902)
Kok, M., Oldenhuis, R., van der Linden, M.P., Meulenber, C.H.,
Kingma, J. and Wiltolt, B.

```

```

TITLE
The Pseudomonas oleovorans alkBAC operon encodes two structurally
related rubredoxins and an aldehyde dehydrogenase
JOURNAL
J. Biol. Chem. 264 (10), 5442-5451 (1989)
MEDLINE
89174582
AUTHORS
3 (bases 1 to 34902)
Eggink,G., Engel,H., Vriend,G., Terstra,P. and Witholt,B.
Rubredoxin reductase of Pseudomonas oleovorans. Structural
relationship to other flavoprotein oxidoreductases based on one NAD
and two FAD fingerprints
JOURNAL
J. Mol. Biol. 212 (1), 135-142 (1990)
MEDLINE
90204534
AUTHORS
4 (bases 1 to 34902)
van Belien,J.B., Eggink,G., Enequist,H., Bos,R. and Witholt,B.
DNA sequence determination and functional characterization of the
OCT-plantlet-encoded alkyl genes of Pseudomonas oleovorans
JOURNAL
Mol. Microbiol. 6 (21), 3121-3136 (1992)
MEDLINE
93086421
AUTHORS
5 (bases 1 to 34902)
Panke,S., Meyer,A., Huber,C.M., Witholt,B. and Mubholts,M.G.
An alkane-responsive expression system for the production of fine
chemicals
JOURNAL
Appl. Environ. Microbiol. 65, 2324-2332 (1999)
AUTHORS
6 (bases 1 to 34902)
van Belien,J.B., Roethlisberger,M. and Witholt,B.
Organization of pseudomonas oleovorans Gpol and Pseudomonas putida
pl genes involved in alkane oxidation
JOURNAL
Unpublished
AUTHORS
7 (bases 1 to 34902)
van Belien,J.B.
Direct Submission
Submitted (21-JUL-1999) van Belien J.B., Institute of
Biotechnology, ETH, ETH-Hoenggerberg, Zuerich CH-8093, Switzerland
Related sequences: AccNo M83932 (IS53), AccNo AF052751 (IS1384),
AccNo U84154 (IS1491).
FEATURES
source
Location/Qualifiers
join(1..15660,16825..34902)
/organism="Pseudomonas putida"
/plasmid="OCT"
/strain="TF4-1L (Gpol)"
/db_xref="taxon:303"
/focus
/note="formerly annotated as Pseudomonas oleovorans TF4-1L
(Gpol)"
CDS
212..751
/note="ORF1"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 35.2 kD"
/protein_id="CAB54044.1"
/db_xref="GI:5824137"
/translation="MIHVAKPRIDPLFNHCHTTPPSIPMLFDSKGFCEWELNSYLS
IGGANGGVPAIKYVFSHADTLNVFOKYTEKKLITLDTDETLYLPAVKYKONSRT
ANSTIKIVRIITALLLEHTQENKFNOLLTTPRAINFQINATEDFYRLNRTPLEF
LTHPCDELRTPTKPTPSFINNELNAMHEAIDHTSNQY1IDRWYTLATLLEHTGSR
IEBAINPASSINAAKNGEVTVNIPVAKGKYGCCLQVLIIPREYIQEESNYSYSLV
AIIYASPMTGYSIAKIIDILMRAPLPITP"
2396..2755
/note="ORF3"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 15.3 kD"
/protein_id="CAB54045.1"
misc_feature
8156 related-like sequence"
CDS
2774..4306
/note="ORF4"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 57.8 kD"
/protein_id="CAB54046.1"
/db_xref="GI:5824139"
/translation="MISMPENLPDDPVLLKOLLROMIHERASDCKITVHLEEVALLR
OILFRKTEQNDATAPOLPLPDEAESAPLDDADDEEYVAPTRKGRKRPADILP
RLEVHELPEHEITCACCGRKHAITEEYSEDLVTPKQIRIVIKHKKVYGRGDESAP
VTADKPAOMIEKSNASPSVLAHLITTKYVDGLPLHREKVLGRGIDIPROTLARWY1
OCSEHFOPLNLMRESLNSRI1HODETRVOYLPEPEKSPSSQSMWVQNGPPDRVY
ILFDYATRAQDEVPRLLDGYRGVMTDYAGYNA1AODGLERLGCMAHARRKFEVA
OKVOPKGTGRADIALNLINKLYGERDLDKSDDEEDLAKVADERSLPLTOLKMWVK
TOPQVTTQNALGAKIAYLANSMSKLERVEHGYLPMONNAEFRA1REFVIGRKWLES
DTPKATASQALYSIVETAKANGCEPYAMLMHALERLPQACSVEDYFALLPMNCTPHI
HS"
4330..4351
/note="identical to right inverted repeat of of IS1131
(1566 family)"
complement(5034..5612)
/note="ORF5"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 21.8 kD"
/protein_id="CAB54047.1"
/db_xref="GI:5824140"
/translation="MVKVSLSPKAPDILLHTEIKWTPPTIWRFRVPENTLGIKH
HIVQVMGSDSHLHEFEIAGEKYGIDPSGNGPVPVPEAKRTLYKALBEKRTFDLY
DFGSMWRIRIKERKLPATRAPOLPYCEGNACPPEDVGVGYGTEGLEAMADPKHP
EHDMMVWGCIPEPAPAFDCERVQWPKRKA"
complement(5940..6230)
/note="ORF6: hypothetical 10.7 kD protein; 60% identity to
putative integrase/recombinase Y4rC, Rhizobium sp. NGR224
(fragment)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB54048.1"
/db_xref="GI:5824141"
/translation="MTLITPSSSRCIAAVWRLISORHRRHATAMHLOAGVDTVIA
LMIGHESPTVTHVEADLAKERFALAIAIAPELRRTYRPPDALLKFLQGL"
complement(6284..7108)
/note="ORF7"
/codon_start=1
/transl_table=1
/transl_except=(pos:complement(7106..7108),aa:Met)
/transl_table=1
/product="hypothetical protein, 31.1 kD"
/protein_id="CAB54049.1"
/db_xref="GI:5824142"
/translation="MISMENKIRMTFRDRLSLHEIAKRTGLARNTIRKVVAREPA
KOPYVORRAIFNKLSTFNAITLFOALDADSLRKQDQRRKALLAIDIKAGYDGSOL
TATIRWKEGCKALQAEVPLSFALGEAFQSEGLIGITVRRMOVAGHICASR
AFMLVAVPSQGHMEFDATRAFSALGVPGRGMPDYVARH1IRLMPFSPAGIGISAA
PRVYGAEEPRGHIRITGNRAISASRPKRLIGMPPRIDVERTGRQPPWFGY"
7188..7214
/note="identical to left inverted repeat of IS53"
7266..7353
/note="similar to IS1384 internal fragment"
7369..7386
/bound_moiety="Alks, positive regulator of alkBFGHKL and
alkST expression"
7424..15611
/evidence="experimental"
/product="alkBFGHKL mRNA"
7520..8725
/gene="alkB"

```

```

CDS
    7520..8725
    /gene="alkB"
    /note="Integral membrane non-heme iron monooxygenase"
    /codon_start=1
    /transl_table=11
    /evidence=experimental
    /product="alkane-1 monooxygenase"
    /protein_id="CAB54050.1"
    /db_xref="GI:5824143"
    /translation="MLEKRVLDSPPEYVDKKYLMISTIMPATPMIGLANETGM
    GIFYGLVLLWGLPLDAMFEDFNPEPEVVKLEKERYITVITLTPMAYAL
    IYAWWGTOPMSWELGALSLGIVNGLANGHEGKKEFDWMMAKIVLAVG
    YGHEFEHNKGRHVDATPMPTSRMGESYIKFSIREIPGAFIRAMGLEORSRIG
    OSVSPFNELLOPMITVITLAVALLFGPKMIVLPIDMGWOLTSANYIEHGL
    LRQKMDGKREHOKPHMSNSNHIVSNLVLPHELRSHHAPRPSYOSLDFGLPA
    LPTGYGATLMAITPMRPSVNDPVMAGDDLKIQLDSSMRITYLKKRTGSAGH
    SSTSAVAS"
    8942..9340
    /gene="alkF"
    8942..9340
    /gene="alkF"
    /codon_start=1
    /transl_table=11
    /product="rubeodioxin 1"
    /protein_id="CAB54051.1"
    /db_xref="GI:5824144"
    /translation="MSRYQCPDQYIYDENKGEPEHGHNTSWNDIFRWACPDCAV

Query Match      64.08; Score 16; DB 3; Length 34902;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 cagcagtcgcatgcc 19
        |||||||||
DB      6204 CAGCAGTGCATGCC 6189

RESULT 27
LOCUS   DMBH48C10 49587 bp DNA INV 04-FEB-2000
DEFINITION Drosophila melanogaster BAC clone BACH48C10.
ACCESSION AL133503
VERSION   AL133503.1 GI:6594136
KEYWORDS
SOURCE
ORGANISM
    fruit fly.
    Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
    1 (bases 1 to 49587)
    Murphy, L., Harris, D., and Barrell, B.
    Sequencing the distal X chromosome of Drosophila melanogaster
    Unpublished
    Sanger Centre, Wellcome Trust Genome Campus, Hinxton Hall, Hinxton,
    Cambridgeshire CB10 1SA, U.K.
    2 (bases 1 to 49587)
    Benos, P.
    Direct Submission
    Submitted (16-DEC-1999) European Drosophila Genome Sequencing
    Consortium
    Clone-BACH48C10; Contig ID=1; Length=49587; Status=Finished
    Sequence submitted by Takis Benos; EMBL Outstation - The EBI,
    Hinxton, Cambridgeshire, CB10 1SD, U.K.
    E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
    Sequencing Consortium. For further information see the European
    Drosophila Genome Sequencing Consortium's web site:
    http://edgp.ebi.ac.uk/.
    The syntax for the representation of annotation used in this record
    is documented at:
    ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation_README
    Coding sequences are predicted from computer analysis, using both
    gene and CDS prediction programs and matches to other sequences.
    These predictions and matches have been evaluated by the annotators

```

```

FEATURES
source
gene
CDS
    and may have been refined by hand (in which case a GeneFinder
    prediction will have no score. The annotators have also used their
    judgement on what matches to represent in this record. A far more
    complete annotation record is available from Flybase
    (http://flybase.bio.indiana.edu/) through the Flybase Annotation
    Object linked by the db_xref qualifier in the feature table.
    IMPORTANT: This sequence is NOT necessarily the entire insert of
    clone BACH48C10. It may be shorter, since we are minimising the
    overlap between clones to 100 bases, by trimming them. Clone
    BACH48C10 overlaps to the left with clone 82C7 the true left end of
    clone BACH48C10 is unknown
    Clone BACH48C10 overlaps to the right with clone BACH7M4 The true
    right end of clone BACH48C10 is unknown
    Sequence in absolute orientation with respect to chromosome.
    Location/Qualifiers
        1..49587
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BAC: BACH48C10"
        /complement(530..1394)
        /gene="EG:BACH48C10.1"
        /complement(join(530..1009,1095..1394))
        /gene="EG:BACH48C10.1"
        /note="prediction=(method:""genefinder"",
        version:""084"", score:""21.51""");
        /prediction=(method:""geneScan"", version:""1.0"",
        score:""73.25""")
        /codon_start=1
        /protein_id="CAB63507.1"
        /db_xref="GI:6594143"
        /translation="MSKKYPNPYTKHFVYSERTKSTCKRCHFDNAGHSHENLAKHLK
        RKHEDAVKSLVDKREIRALAAQOANKRANVAPTRKVLASKRPLDPTPKOSFTIEDD
        DSOHSGKEMWTNCGEETGVDVKNKATVDFEGVIGKSEIPDEYQKMSANNEIK
        AVDSVAQITNESSVCYAPRTSTTARAKSPARASSAAALDOSPDEDFLRLTSGKLG
        KYSARNTVQVFIHRLTKADMGHEFTDSMGDDVYSP"
        2911..4557
        /gene="EG:BACH48C10.2"
        /join(2911..3312,3484..4557)
        /gene="EG:BACH48C10.2"
        /note="prediction=(method:""genefinder"",
        version:""084"", score:""52.65""");
        /prediction=(method:""geneScan"", version:""1.0"",
        score:""112.75"""); /match=(desc:""GH19593.5prime GH
        Drosophila melanogaster head port Drosophila melanogaster
        CDNA clone GH19593 5prime, mRNA sequence""",
        species:""Drosophila melanogaster (fruit fly)"",
        ranges:(query:2855..3324, target:EMBL::A138635.1..470,
        score:""2296.00"""), (query:3478..3538,
        target:EMBL::A138635.453..513, score:""287.00"""),
        method:""blastn"", version:""1.4.9""")
        /codon_start=1
        /protein_id="CAB63508.1"
        /db_xref="GI:6594144"
        /translation="WTLRRLTPRRIVDYCRRLSTDVYATLAPFYQAIPTQGIANP
        AKRNVESLRLCLSHLESILPSPASCLPDGHLEFISGSDSSRYVSRVRYQV
        DRLIDPRAVTCQLIRSCKLPRKEQEQMKRQDEKELGVLRKEPVREKRPQV
        PVOKRLALRELFPDKAREHSLIEMRQALARSQASVATLRKIGITAEALVOLLGI
        RGTNRDCCNMERNYHNCISILKISGDEPKYAGQALARNFRVFAVYTGISLEGDVL
        FSGDSPSKWLAFLSEMRLHEORLRLVPLDKALSAVLRIGIRERRRVPDEGCARAV
        GSLRVIRAVSEHRLIEDSVDELPAASLDHHLIATVPDSIPIKVSQGLFLAPASCSP
        DLVEFICNLDVAVAGRSRYRRDQVRQMLRQCCDELGIGRLSKDSDVTPDKMMAI
        RHLDSGVKSCRGSLHTNTYNSVPTGIVCIPDMSRER"
        16590..19135
        /gene="EG:BACH48C10.3"
        /join(16590..17843,17908..18048,18117..18490,18556..19135)
        /gene="EG:BACH48C10.3"
        /note="The EST emb1AI063630AI063630 comes from the 5'
        UTR"
        /codon_start=1
        /protein_id="CAB72239.1"
        /db_xref="GI:6911910"
        /translation="MSRSSTGREGSDLYDPLAEHLHWOLVKHTRENIDALNAKFEAN
        LOEPAMVLIIGSKAEILNTTWLGNPTPTSKLPIKITYVYVVISLMLPLPEYGOEL

```



```

TSKLEIAKEEDELMEERLNSODDOEDSSLVFERKPHYQNOTOILQOORLARVHHG
NDLTPDSQSPSSCGCTLTROKILLRAHLBPQOQTSVEISGVRLCDALMKALKRQ
LTPDMCEVSTHSGKHIIIPWHTDIGTLHVEEIPALIDKPEIRTHIKOIIIRKTFPSL
VEBCCRLLFTGFCYCSQCNFRHORCANRYMLCQPPMDSYOLLAEHPDNGVGP
PGRGAVFRNMSRSRSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
VCINIRSVTSEVORSLIMQARPLPHCTGHSNSTQSPSTTKHNHRRRSDSPN
KLLLRDAXSSEENNTILAEFLIGRPGTSCFCTVRAHHGGVAYKTILAVKTFEPA
QLOAFKNEVAMUKTKRHGNTILLFMCVCSPSLAYTOWCESSSLYKHHVSETEKLN
TLIDGROVAGMOTYLAKNIIIRHDKSNNTLEHEDSVKIGDGLTAKTRMGEO
ANOPTGSLIMAEVIRMOETINPVSFQSDYAFGIIVMEYELAECLPYGHIISNKQIIF
WVGRLLPMDSDVQSDAPQALRIAEEDICIKYTPKDRPLMLNMLNMLRTLPKIH
RSASEPNLTQSOILDEFILPSPKTPVFNFNQFPGSAGNI"
complement(19754..20239)
/gene="EG:BACH48C10.6"
complement(19754..20239)
/gene="EG:BACH48C10.6"
complement(19754..20239)
/gene="EG:BACH48C10.6"
version="084""; score="26.59"";
/prediction=(method:"genscan", version:"1.0"";
score:"82.07""; /match=(desc:"SIMILARITY TO RIBOSOMAL
PROTEIN L14", species:"Caenorhabditis elegans",
ranges:(query:20027..20128, target:SPTRMBL::020430:36..3,
score:"94.00""), (query:19952..20011,
target:SPTRMBL::020430:67..48, score:"73.00""),
(query:19835..19924, target:SPTRMBL::020430:109..80,
score:"88.00""), method:"blastx", version:"1.4.9""")
/codon.start=1
/protein_id="CAB63504.1"
/db_xref="GI:6594140"
/translation="MALRLIKTVRQVAPLAVLTGQGOQTOLHTTPACCEIRKLRL
RVVNSDLGRKAMAEGRPCIHVYKRGVIGDYVLAIKCMKGLVGLKONOK
PKQKPFDSNNLVLIIDNGSPSLGTRHPIPTILRLIKEKTLAKGADYTKVLAASRY
V"
complement(21238..25103)
/gene="EG:BACH48C10.5"
complement(21238..25103)
/gene="EG:BACH48C10.5"
/ncle="/prediction=(method:"genscan", version:"1.0"";
score:"314.45""; /prediction=(method:"genefinder",
version:"084""; /match=(desc:"GH02860 5prime GH
Drosophila melanogaster head port2 Drosophila melanogaster
cDNA clone GH02860 5prime, mRNA sequence",
species:"Drosophila melanogaster (fruit fly)",
ranges:(query:25045..25290,
target:EMBL::A1064644:369..124, score:"1221.00""),
(query:24786..24952, target:EMBL::A1064644:535..369,
score:"817.00""), method:"blastn", version:"1.4.9"";
/match=(desc:"LD12603 5prime LD Drosophila melanogaster
embryo Bluescript Drosophila melanogaster cDNA clone
LD12603 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:25040..25317,
target:EMBL::AA438345:305..28, score:"1363.00""),
(query:24647..24952, target:EMBL::AA438345:604..229,
score:"1422.00""), (query:24549..24579,
target:EMBL::AA438345:635..605, score:"155.00""),
method:"blastn", version:"1.4.9"";
/match=(desc:"bs12f02.yl Drosophila melanogaster adult
testis library Drosophila melanogaster cDNA clone bs12f02
5', mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:23494..24011,
target:EMBL::AF945413:522..5, score:"2554.00""),
method:"blastn", version:"1.4.9"";
/match=(desc:"GM04080 5prime GM Drosophila melanogaster
ovary Bluescript Drosophila melanogaster cDNA clone
GM04080 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:22290..22894,
target:EMBL::AA802249:605..1, score:"2971.00""),
method:"blastn", version:"1.4.9"";
/match=(desc:"GH25306 5prime GH Drosophila melanogaster
head port2 Drosophila melanogaster cDNA clone GH25306
5prime, mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:21634..22176,
target:EMBL::AI405337:543..1, score:"2706.00""),
method:"blastn", version:"1.4.9""")
/codon.start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMGSFATLNRMGRRKDKSGTGNSSMGKLSKSTOLHLST
DTDINISOLEVRIIPVPAATSNAPEDQTFRTVYLLPKDQLEFVARIGARVLSKLS
LVCDNKLQADAEKFRSPYDASQVYCESTIGVGLSELRLCHKSSSYDNFPMYKLS
FQRTGVARDSILSSSDPSSSRHSHKAKRPYSSNSNLSGSPDLDEVLDTITIN
AKVLAAPPKRRKRPAPPPOVCIIPROSVVDIIRANKSPSPATAYIVYKROLCMSPNL
PQPSQSDNISTASLEPEVYTPPNTIPAPRTTTPLPAPVANPLAONNGNSPN
GNASKVLNRTPPEPERSLGSATVEDDNDNTASQKODSGEPAEPSVSPPEAKSO
QESSSEDDDAIKVYNFKLCQTSKKPEPSAATLAEITLTAODEEDNHVDQPAITN
GNATPSSSETSLSNMYLDPISPDPSPQOQKRNAGLSGSPDLDEVLDTITIN
OORLDTLIRKOPDPQWVEIESAKPNMLAFSTTTAKTTRIGRADSPRQAGAGGTEV
ASSGRALSHRSSSHVSLNKLBNQICIGORSSSELSTIGESSLQSDPEVITILSRK
NSIAESGASAPVYBQPKITQULSLAKSSPAVQKQSDISSRSEASQOSTPVPK
SNQKVEKPSPTFRVQDEPSPSPALQKSSPAVQKQSDISSRSEASQOSTPVPK
SLHQKIVQKESPLESSKESFVYSKMSPPAVKVPDPPEPKVAVEESSGNSSSO
STTPNPPTATPVSPVNLNPNMTVAIEIOSQORDPAPPSKPLPTTYRSGPSPINFA
method:"blastn", version:"1.4.9""")
/codon.start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMGSFATLNRMGRRKDKSGTGNSSMGKLSKSTOLHLST
DTDINISOLEVRIIPVPAATSNAPEDQTFRTVYLLPKDQLEFVARIGARVLSKLS
LVCDNKLQADAEKFRSPYDASQVYCESTIGVGLSELRLCHKSSSYDNFPMYKLS
FQRTGVARDSILSSSDPSSSRHSHKAKRPYSSNSNLSGSPDLDEVLDTITIN
AKVLAAPPKRRKRPAPPPOVCIIPROSVVDIIRANKSPSPATAYIVYKROLCMSPNL
PQPSQSDNISTASLEPEVYTPPNTIPAPRTTTPLPAPVANPLAONNGNSPN
GNASKVLNRTPPEPERSLGSATVEDDNDNTASQKODSGEPAEPSVSPPEAKSO
QESSSEDDDAIKVYNFKLCQTSKKPEPSAATLAEITLTAODEEDNHVDQPAITN
GNATPSSSETSLSNMYLDPISPDPSPQOQKRNAGLSGSPDLDEVLDTITIN
OORLDTLIRKOPDPQWVEIESAKPNMLAFSTTTAKTTRIGRADSPRQAGAGGTEV
ASSGRALSHRSSSHVSLNKLBNQICIGORSSSELSTIGESSLQSDPEVITILSRK
NSIAESGASAPVYBQPKITQULSLAKSSPAVQKQSDISSRSEASQOSTPVPK
SNQKVEKPSPTFRVQDEPSPSPALQKSSPAVQKQSDISSRSEASQOSTPVPK
SLHQKIVQKESPLESSKESFVYSKMSPPAVKVPDPPEPKVAVEESSGNSSSO
STTPNPPTATPVSPVNLNPNMTVAIEIOSQORDPAPPSKPLPTTYRSGPSPINFA

```

```

Query Match          64.0%; Score 16; DB 6; Length 49587;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcattgccagcatt 25
|||||
Db 27002 GTGCATGCCAGCATTT 26987

RESULT 29
ATAC009895

method:"blastn", version:"1.4.9""")
/codon.start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMGSFATLNRMGRRKDKSGTGNSSMGKLSKSTOLHLST
DTDINISOLEVRIIPVPAATSNAPEDQTFRTVYLLPKDQLEFVARIGARVLSKLS
LVCDNKLQADAEKFRSPYDASQVYCESTIGVGLSELRLCHKSSSYDNFPMYKLS
FQRTGVARDSILSSSDPSSSRHSHKAKRPYSSNSNLSGSPDLDEVLDTITIN
AKVLAAPPKRRKRPAPPPOVCIIPROSVVDIIRANKSPSPATAYIVYKROLCMSPNL
PQPSQSDNISTASLEPEVYTPPNTIPAPRTTTPLPAPVANPLAONNGNSPN
GNASKVLNRTPPEPERSLGSATVEDDNDNTASQKODSGEPAEPSVSPPEAKSO
QESSSEDDDAIKVYNFKLCQTSKKPEPSAATLAEITLTAODEEDNHVDQPAITN
GNATPSSSETSLSNMYLDPISPDPSPQOQKRNAGLSGSPDLDEVLDTITIN
OORLDTLIRKOPDPQWVEIESAKPNMLAFSTTTAKTTRIGRADSPRQAGAGGTEV
ASSGRALSHRSSSHVSLNKLBNQICIGORSSSELSTIGESSLQSDPEVITILSRK
NSIAESGASAPVYBQPKITQULSLAKSSPAVQKQSDISSRSEASQOSTPVPK
SNQKVEKPSPTFRVQDEPSPSPALQKSSPAVQKQSDISSRSEASQOSTPVPK
SLHQKIVQKESPLESSKESFVYSKMSPPAVKVPDPPEPKVAVEESSGNSSSO
STTPNPPTATPVSPVNLNPNMTVAIEIOSQORDPAPPSKPLPTTYRSGPSPINFA

```

```

RESULT 28
AC018267/c
LOCUS AC018267
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC018267
VERSION AC018267.1 GI:6552924
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 60573)
REFERENCE Adams,M. and Venter,J.C.
AUTHORS Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214125 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..60573
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 16964 a 14192 c 13649 g 15768 t
ORIGIN

```

```

Query Match          64.0%; Score 16; DB 6; Length 60573;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcattgccagcatt 25
|||||
Db 27002 GTGCATGCCAGCATTT 26987

RESULT 29
ATAC009895

```


LOCUS ATAC009895 82289 bp DNA PLN 24-JAN-2001
 DEFINITION Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,
 complete sequence.
 ACCESSION AC009895
 VERSION AC009895.4 GI:12408719
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 82289)
 AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
 Rong, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
 Bowman, C.L., White, O., Nieman, M.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence
 unpublished
 TITLE 2 (bases 1 to 82289)
 AUTHORS Lin, X. and Kaul, S.
 REFERENCE Direct Submission
 TITLE Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
 JOURNAL Medical Center Dr., Rockville, MD 20850, USA, xlinetlgr.org
 3 (bases 1 to 82289)
 REFERENCE Lin, X.
 AUTHORS Direct Submission
 TITLE Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
 JOURNAL Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280866.
 COMMENT Address all correspondence to:
 Xiaoyang Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlinetlgr.org
 BAC clone T21P5 is from Arabidopsis chromosome III and is near the
 molecular marker ml172.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
 Washington), Genscan (Chris Burge,
 http://genome.stanford.edu/~chris/GENSCAN.html), and NeplantGene
 (http://www.cds.dtu.dk/neplantgene/cdsneplantgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tldb/at/at.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted as 'trnscan-SE' (Sean Eddy,
 http://genome.wustl.edu/eddy/Trnscan-SE/). Simple repeats are
 identified by repeatmasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRAIL are annotated as misc features.
 location/Qualifiers
 1..82289
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="ml172"
 /clone="T21P5"
 1..24217
 /note="Overlap with BAC clone T12T13
 (AC009327:65718..89934)."
 complement(join(<8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..>3100))

gene /gene="T21P5.1"
 complement(<8..>3100)
 /gene="T21P5.1"
 /note="similar to hypothetical protein GB:CA38918
 [Arabidopsis thaliana]"
 complement(join(<8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..3100))
 CDS /gene="T21P5.1"
 /note="unknown protein"
 /codon_start=1
 /protein_id="AAAF01600.1"
 /db_xref="GI:6017117"
 /translation="MSQDISGNNNPISSEVYWSLVNKADEKFKIRDLPEPERSRYEN
 YPKPKVYTOQMKROENRQKIVAGIKRMEIGISRIQOLYGVHMYRTSDAGYLS
 ESYVEFALITRETKFDGLPDNLNANKQLFLARFLRMVLCVIGRRMVDVQDFRS
 LIDCKRFQETDERKEMVVADEIVFLKSDTAPRNPPLPSYVDLPNADGPRRS
 RSLRLTDALTSYCNKVEVYSELTLDSFRMLQCLEWPSGSYIYSDLGAKGNAPVY
 ARINSQSNMPTLPNPKRAVLYRPSITHFLAVLATICEELSHGILLIYLSAGKTG
 OTSSPSLSARSATSVSEENILRDESHITKQETESLQITPSSGSYIYSDLVPTTRP
 LPTIDSSYTFYFNKICGAKGEPALLDSHPPLISADFSRQPSGLFTIFLTPSP
 VQAFCLSDTSMDMETDITFYAEKLLSSNMWASLTATSDTLHPWVSQILKQPFIR
 RLRLRTFCRAVALTTFVFNKQNPQECSTLPSLLPTAPNQSVAFTQMANVFGAT
 SKETIPQDITMDESP"
 join(<3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..56144)
 /gene="T21P5.2"
 <3627..>6144
 /gene="T21P5.2"
 /note="predicted by genscan"
 join(3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..6144)
 /gene="T21P5.2"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="AAAF01580.1"
 /db_xref="GI:6017097"
 /translation="MDNRSSESIRKHEIEKPTIASRKLEDTNFKLIQDPPEMALYAK
 VRSEPEIHSLOIRIAACIKMDKMLNEKYEGLERKCAQDRAVIDEKONESYSANEL
 ARRGQDLEMLKLAHDLKTYEDRPRTYFMSITGLKLAEGVWMPRNANATISGIIHL
 DQLOMKTRKACNDRIRELSIVENQPTGTISKDNHDPNRKQASGICSDRENQDTN
 EOLLPMENYTRNPNYHIMQDTELSFNNQIGGSGGICPQKRENFGLPSSVAKGE
 MIOREKKAENSSMFDAYNGNEEFASHVEEGGIDGFOIIDLALPEGVVLCGPVR
 GTTICMOWNRHLEDGROYIEGATPEYIVTADVDKLIAVECICPMDOGRQVYRD
 FSGIYSNESVSKDVLINQGLVRLFANDQKIRCGVFNQGT"
 6225..6289
 /note="exon predicted by xgrail, quality marginal"
 6384..6503
 /note="exon predicted by xgrail, quality excellent"
 6794..6839
 /note="exon predicted by xgrail, quality excellent"
 6928..7033
 /note="exon predicted by xgrail, quality marginal"
 7293..7411
 /rpt_family="(TAAA)n"
 complement(7488..7547)
 /note="exon predicted by xgrail, quality
 excellent_shadowexon"
 complement(7654..7784)
 /rpt_family="(TAA)n"
 complement(7819..7864)
 /note="exon predicted by xgrail, quality good_shadowexon"
 8054..8113
 /note="exon predicted by xgrail, quality marginal"
 8263..8380
 /rpt_family="(TAAA)n"
 complement(8592..8627)
 /note="exon predicted by xgrail, quality
 marginal_shadowexon"
 8652..>9527
 /gene="T21P5.3"

```

gene
    <8652..>9527
    /gene="T21P5.3"
    /note="contains zinc finger motif, C3HC4 type (RING
    finger)"
    8652..9527
    /gene="T21P5.3"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AA01602.1"
    /db_xref="GI:6017119"
    /translation="MSGTGNPNPWTYDSYRDCSGQVCVYCPQWCYIPEPPSPF
    LDDDESSSDSPFLIALIGLISAFILVSYTILSKYCHRRNRSSTSAANRI
    SDYTQGTNNNGATINNOTIGAGGDESLIKITVYKKRMKQFESSDCS
    VCLSEFENESLRLPKCNAPFVPCIDIMKLSHNCPCRAPIYSSAVELVINO
    OIYENNSISTGDSVVVNLDELNRSKNEYVNESTPTXGDAGFKRRRAETGVK
    LRRCGIDRYRTGCD"
misc_feature
    complement(10838..10896)
    /note="exon predicted by xgrail, quality
    marginal_shadowexon"
misc_feature
    11174..11254
    /note="exon predicted by xgrail, quality good"
misc_feature
    11690..11700
    /note="exon predicted by xgrail, quality marginal"
mRNA
    join(<11814..12256,12381..12578,12676..13085,13162..13332,
    13418..>13761)
    /gene="T21P5.4"
    <11814..>13761
    /gene="T21P5.4"
    /note="similar to unknown protein GB:AA032238 [Arabidopsis
    thaliana]"
    join(11814..12256,12381..12578,12676..13085,13162..13332,
    13418..13761)
    /gene="T21P5.4"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AA01581.1"
    /db_xref="GI:6017098"
    /translation="MAETKKGSESPITITVVLVOENRFDHTLGWKEKLEIDGYM
    KSDOKFNEFTSDLSHNVFEDQSYVDPPNGHSIRDIYEOVGKPMDSGPPNG
    PATMSGFAMNARKKMSAVMNGKFDALPYVKALONFALIDRFASVPGAIQON
    RYLIHSATSHGTNNERKLLIEGPOKTIPESLDEAGTFTGYOCPTTLFPRNLK
    LKYLTFHDYGLQFKCKEKNLPPYVVGQRMVDLLNPNDDPSPHDSSECKLYK
    EYFALRSSPQNNELLFITITDEHGGFVDPPLDCCVNDGILGPPINFEENRG
    VVPEFTISPMTEPGTVHSGNCPYIMSQYHSSIPATVKIRLKDPLTRDSMAGT
    FESVITRNSPQDCETILSNVYKMGATVAKENAEISDFBELVIVAAGLDYKNEEL
    LYKICKTCVDSASKTYTKAFDKFVEESKKAREGGDNDIVFCVDDDDHNVKPPP
    SOSPSHATPWSN"
misc_feature
    complement(14488..14563)
    /note="exon predicted by xgrail, quality
    excellent_shadowexon"
misc_feature
    14893..14935
    /note="exon predicted by xgrail, quality
    marginal"
    join(<15141..15386,15675..15872,16258..16667,16780..16950,
    17036..>17427)
Query Match
    Best Local Similarity 100.0%; Pred. No. 68;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy 8 agtgatgcgcagca 23
    Db 30302 AGTGTGATGCCGCA 30317
RESULT 30
AC011732 115955 bp DNA HTG 14-MAR-2000
LOCUS Homo sapiens clone RP11-17011, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC011732
AC011732.4 GI:7107960
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human,

```

```

ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-17011
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 115955)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,L., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearlano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,M., Forrest,C., Funke,R., Gage,D.,
Galligan,J., Gardyna,S., Grant,G., Haas,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
McEwen,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 28, 2000 this sequence version replaced gi:6524206.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 17.0.11
Center clone name: 13676
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 91165 bases at least Q40
Consensus quality: 102606 bases at least Q30
Consensus quality: 109983 bases at least Q20
Insert size: 87000; agarose-fp
Insert size: 114655; sum-of-contigs
Quality coverage: 6.0 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1835: contig of 1835 bp in length
* 1836 1935: gap of 100 bp
* 1936 3722: contig of 1787 bp in length
* 3723 3822: gap of 100 bp
* 3823 6826: contig of 3004 bp in length
* 6827 6926: gap of 100 bp
* 6927 12743: contig of 5817 bp in length
* 12744 12843: gap of 100 bp
* 12844 19565: contig of 6722 bp in length
* 19566 19665: gap of 100 bp
* 19666 25342: contig of 5677 bp in length
* 25343 25442: gap of 100 bp
* 25443 35236: contig of 9794 bp in length
* 35237 35336: gap of 100 bp
* 35337 45543: contig of 10207 bp in length
* 45544 45643: gap of 100 bp

```

```

* 45644 57815: contig of 12172 bp in length
* 57816 57915: gap of 100 bp
* 57916 73012: contig of 15097 bp in length
* 73013 73112: gap of 100 bp
* 73113 87110: contig of 13998 bp in length
* 87111 87210: gap of 100 bp
* 87211 95720: contig of 8510 bp in length
* 95721 95820: gap of 100 bp
* 95821 102140: contig of 6320 bp in length
* 102141 102240: gap of 100 bp
* 102241 115955: contig of 13715 bp in length.
Location/Qualifiers
1. 115955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-17011"
/clone_lib="RPCT-11 Human Male BAC"

```

```

misc_feature 1. 1835
/note="assembly_fragment"
1936. 3722
/note="assembly_fragment"
clone_end:T7
vector_side:left"
misc_feature 3823. 6826
/note="assembly_fragment"
6927. 12743
/note="assembly_fragment"
12844. 19565
/note="assembly_fragment"
19666. 25342
/note="assembly_fragment"
25443. 35236
/note="assembly_fragment"
35337. 45543
/note="assembly_fragment"
45644. 57815
/note="assembly_fragment"
57916. 73012
/note="assembly_fragment"
73113. 87110
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature 87211. 95720
/note="assembly_fragment"
95821. 102140
/note="assembly_fragment"
102241. 115955
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
BASE COUNT 31603 a 25627 c 26285 g 31102 t 1338 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 62; Length 115955;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 3 gcaagcaggtgcatcgc 18
|||||
Db 106529 GCACGACGTCATCGC 106544

```

```

RESULT 31
AC002457/c AC002457 177380 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone CTB-60P12 from Tq21, complete sequence.
DEFINITION AC002457
ACCESSION AC002457
VERSION AC002457.1 GI:3947433
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

```

REFERENCE 1 (bases 1 to 177380)
AUTHORS Smith,A., Beck,C. and Gibson,A.
TITLE The sequence of Homo sapiens BAC clone CTB-60P12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 3 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Dec 2, 1998 this sequence version replaced g1:2337872.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@wuston.wustl.edu
----- Summary Statistics
Center project name: H_RG060P12

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
Clone CTB-60P12 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTB-60N22, 200 bp overlap; the clone sequenced to the right is CTB-137N13, 200 bp overlap. Actual start of this clone is at base position 1 of CTB-60P12; actual end is at 177380 of CTB-60P12.

FEATURES

source

```

1. 177380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21"
/clone="CTB-60P12"

```

```

gene      /clone_lib="CTP-978SK-B"
           complement(<131..112890)
           /gene="WUGSC:H_RG060P12.2"
           complement(join(<131..215,1834..1923,31059..31144,
41160..41293,90777..90892,112769..112890))
CDS        /gene="WUGSC:H_RG060P12.2"
           /note="Rapid interacting protein; similar to U73941
(PID:q1916018); H_RG060P12.2"
           /codon_start=1
           /evidence=not_experimental
           /protein_id="AAC82532.1"
           /db_xref="GI:3947434"
           translation="MASRSLGSLRGCGGCGCKSLSRNNAVERNNLITGRSV
KVLIDRSCFEETIDSSPEPNPAILLEOLISHLKQGVTEGESPEFMDYTRVACR
IVGSEANMLAGMILGLNIDFSCFKGEGLDGSPFAVIDYTPYLKIQY"
repeat_region 480..791
             /rpt_family="Alu"
repeat_region 4518..4824
             /rpt_family="Alu"
repeat_region 4825..4937
             /rpt_family="(TA)n"
repeat_region 5732..6004
             /rpt_family="L2"
repeat_region 6012..6047
             /rpt_family="L2"
repeat_region 6143..6231
             /rpt_family="MIR"
repeat_region 6232..7254
             /rpt_family="L2"
repeat_region 7272..7412
             /rpt_family="L1"
repeat_region 7420..7813
             /rpt_family="L1"
repeat_region 7814..8150
             /rpt_family="MIR"
repeat_region 8151..8265
             /rpt_family="MER1_type"
repeat_region 8294..9183
             /rpt_family="L1"
repeat_region 9314..10441
             /rpt_family="Retroviral"
repeat_region 10437..12545
             /rpt_family="L1"
repeat_region 12928..13352
             /rpt_family="L1"
repeat_region 14327..14443
             /rpt_family="MaLR"
repeat_region 14443..16612
             /rpt_family="L1"
repeat_region 16719..16804
             /rpt_family="L1"
repeat_region 17093..17116
             /rpt_family="AT_rich"
repeat_region 17492..17774
             /rpt_family="AT_rich"
repeat_region 18097..18135
             /rpt_family="AT_rich"
repeat_region 19255..19594
             /rpt_family="L1"
repeat_region 19900..19946
             /rpt_family="L1"
repeat_region 19966..20245
             /rpt_family="AT_rich"
repeat_region 20549..20614
             /rpt_family="L1"
repeat_region 20684..20816
             /rpt_family="AT_rich"
repeat_region 20817..21123
             /rpt_family="AT_rich"
repeat_region 21124..21134
             /rpt_family="Alu"
repeat_region 21136..21171
             /rpt_family="AT_rich"

```

```

repeat_region 21478..21940
             /rpt_family="(CA)n"
repeat_region 22310..23003
             /rpt_family="L1"
repeat_region 23135..23653
             /rpt_family="L1"
repeat_region 23658..24140
             /rpt_family="MER21_g"
repeat_region 24263..24304
             /rpt_family="L1"
repeat_region 24419..24447
             /rpt_family="AT_rich"
repeat_region 24618..24865
             /rpt_family="AT_rich"
repeat_region 25627..25650
             /rpt_family="Alu"
repeat_region 25906..26152
             /rpt_family="(TTAA)n"
repeat_region 26153..26536
             /rpt_family="L2"
repeat_region 26654..27556
             /rpt_family="AT_rich"
repeat_region 28842..28903
             /rpt_family="Retroviral"
repeat_region 29888..30080
             /rpt_family="(CAT)n"
repeat_region 30382..30669
             /rpt_family="MER1_type"
repeat_region 32790..32960
             /rpt_family="Alu"
repeat_region 33062..33192
             /rpt_family="MIR"
repeat_region 33199..33398
             /rpt_family="L1"
repeat_region 34000..34042
             /rpt_family="L1"
repeat_region 34043..34316
             /rpt_family="(TAAA)n"
repeat_region 34414..34602
             /rpt_family="Alu"
repeat_region 35064..35263
             /rpt_family="L1"
repeat_region 36255..36571
             /rpt_family="L1"
repeat_region 36571..36571
             /rpt_family="Alu"

Query Match      64.0%; Score 16; DB 85; Length 177380;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      8 aggtgcatcgccagca 23
DB      48179 AGGTGCAATCGCCAGCA 48164

RESULT 32
AC084136/c 180700 bp DNA HTG 14-FEB-2001
LOCUS      AC084136
DEFINITION Homo sapiens clone RP13-506N4, WORKING DRAFT SEQUENCE, 32 unordered
            pieces.
ACCESSION  AC084136
VERSION    AC084136.2
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 180700)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens clone RP13-506N4
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 180700)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

TITLE
JOURNAL
COMMENT

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
Boukarguier, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliiev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K.,
Lamarez, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKenna, K.,
McPheters, R., Meldrim, J., Menes, L., Mihova, T., Mianga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, T.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Ribbeck, M., Riley, R.,
Rogov, P., Rottman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirelli, A., Travers, M., Trifillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 14, 2001 this sequence version replaced g1:10800269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L11406

Center Clone name: 506_N_4

Summary Statistics

Sequencing vector: plasmid; N/A; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 167122 bases at least Q40

Consensus quality: 174060 bases at least Q30

Consensus quality: 176531 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 177600; sum-of-coverage

Quality coverage: 3.7 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 32 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 951: contig of 951 bp in length
* 952 1051: gap of 100 bp
* 1052 2064: contig of 1013 bp in length
* 2065 2164: gap of 100 bp
* 2165 3416: contig of 1252 bp in length
* 3417 3516: gap of 100 bp
* 3517 4816: contig of 1300 bp in length
* 4817 4916: gap of 100 bp
* 4917 6297: contig of 1381 bp in length
* 6298 6397: gap of 100 bp
* 6398 9144: contig of 2747 bp in length
* 9145 9244: gap of 100 bp
* 9245 11578: contig of 2334 bp in length
* 11579 11678: gap of 100 bp
* 11679 30905: contig of 19227 bp in length
* 30906 31005: gap of 100 bp
* 31006 32606: contig of 1601 bp in length
* 32607 32706: gap of 100 bp
* 32707 34256: contig of 1550 bp in length

FEATURES
source

34257 34356: gap of 100 bp
* 34357 37618: contig of 3262 bp in length
* 37619 37718: gap of 100 bp
* 37719 39131: contig of 1413 bp in length
* 39132 39231: gap of 100 bp
* 39232 41291: contig of 2060 bp in length
* 41292 41391: gap of 100 bp
* 41392 43172: contig of 1781 bp in length
* 43173 43272: gap of 100 bp
* 43273 45866: contig of 2614 bp in length
* 45867 45986: gap of 100 bp
* 45987 49186: contig of 3200 bp in length
* 49187 49286: gap of 100 bp
* 49287 52781: contig of 3495 bp in length
* 52782 52881: gap of 100 bp
* 52882 57288: contig of 4407 bp in length
* 57289 57388: gap of 100 bp
* 57389 62317: contig of 4929 bp in length
* 62318 62417: gap of 100 bp
* 62418 65388: contig of 2971 bp in length
* 65389 65488: gap of 100 bp
* 65489 70061: contig of 4573 bp in length
* 70062 70161: gap of 100 bp
* 70162 75771: contig of 5610 bp in length
* 75772 75871: gap of 100 bp
* 75872 83423: contig of 7552 bp in length
* 83424 83523: gap of 100 bp
* 83524 89716: contig of 6193 bp in length
* 89717 89816: gap of 100 bp
* 89817 97026: contig of 7210 bp in length
* 97027 97126: gap of 100 bp
* 97127 104654: contig of 7528 bp in length
* 104655 104754: gap of 100 bp
* 104755 114340: contig of 9586 bp in length
* 114341 114440: gap of 100 bp
* 114441 126202: contig of 11762 bp in length
* 126203 126302: gap of 100 bp
* 126303 142208: contig of 15906 bp in length
* 142209 142308: gap of 100 bp
* 142309 159777: contig of 17469 bp in length
* 159778 159877: gap of 100 bp
* 159878 178832: contig of 18855 bp in length
* 178833 178932: gap of 100 bp
* 178933 180700: contig of 1768 bp in length.
Location/Qualifiers

FEATURES

1..180700

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP13-506N4"

/clone_lib="RP13-13 Human Female BAC 481-768"

1..951

/note="assembly-fragment"

/clone_end:SP6

/vector_side:left"

misc_feature

1052..2064

/note="assembly-fragment"

2165..3416

/note="assembly-fragment"

3517..4816

/note="assembly-fragment"

4917..6297

/note="assembly-fragment"

6398..9144

/note="assembly-fragment"

9245..11578

/note="assembly-fragment"

11679..30905

/note="assembly-fragment"

31006..32606

/note="assembly-fragment"

32707..34256

/note="assembly-fragment"

34357..37618

misc_feature

Query Match 60.0%; Score 15; DB 5; Length 3069;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 agcaggtcgcgcgc 19
 |||
 Db 2114 AGCAGGTGCGCCGC 2100

RESULT 37
 AF274670 3369 bp DNA PLN 13-AUG-2000
 LOCUS AF274670/c
 DEFINITION Tetracentron sinense 26S ribosomal RNA gene, complete sequence.
 ACCESSION AF274670
 VERSION AF274670.1 GI:9799468
 KEYWORDS
 SOURCE Tetracentron sinense.
 ORGANISM Tetracentron sinense.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Trochodendraceae; Tetracentroideae;
 Tetracentron.

REFERENCE 1 (bases 1 to 3369)
 Fishbein,M., Hibsch-jetter,C., Soltis,D.E. and Hufford,L.
 Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
 Rapid, Ancient Radiation
 Unpublished
 2 (bases 1 to 3369)
 Fishbein,M., Hibsch-jetter,C., Soltis,D.E. and Hufford,L.
 Direct Submission
 Submitted (30-MAY-2000) School of Biological Sciences, Washington
 State University, Pullman, WA 99164-4236, USA

FEATURES
 source
 1..3369
 /organism="Tetracentron sinense"
 /specimen_voucher="Qiu 94166 (IND)"
 /db_xref="taxon:13715"
 1..3369
 /product="26S ribosomal RNA"
 BASE COUNT 806 a 824 c 1049 g 689 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 3369;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcgcgcgcgc 22
 |||
 Db 441 AGGTGCATCGCCAGC 427

RESULT 38
 AF274671 3369 bp DNA PLN 13-AUG-2000
 LOCUS AF274671/c
 DEFINITION Trochodendron aralioides 26S ribosomal RNA gene, complete sequence.
 ACCESSION AF274671
 VERSION AF274671.1 GI:9799469
 KEYWORDS
 SOURCE Trochodendron aralioides.
 ORGANISM Trochodendron aralioides.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Trochodendraceae; Trochodendroideae;
 Trochodendron.

REFERENCE 1 (bases 1 to 3369)
 Fishbein,M., Hibsch-jetter,C., Soltis,D.E. and Hufford,L.
 Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
 Rapid, Ancient Radiation
 Unpublished
 2 (bases 1 to 3369)
 Fishbein,M., Hibsch-jetter,C., Soltis,D.E. and Hufford,L.
 Direct Submission
 Submitted (30-MAY-2000) School of Biological Sciences, Washington
 State University, Pullman, WA 99164-4236, USA

FEATURES
 source
 Location/Qualifiers
 1..3369
 /organism="Trochodendron aralioides"
 /specimen_voucher="Qiu 94157 (IND)"
 /db_xref="taxon:4407"
 1..3369
 /product="26S ribosomal RNA"
 BASE COUNT 801 a 816 c 1060 g 690 t 2 others
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 3369;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcgcgcgcgc 22
 |||
 Db 441 AGGTGCATCGCCAGC 427

RESULT 39
 MTCY9C4 15916 bp DNA BCT 17-JUN-1998
 LOCUS MTCY9C4
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 113/162.
 ACCESSION Z77250 AL123456
 VERSION Z77250.1 GI:3261617
 KEYWORDS
 SOURCE Mycobacterium tuberculosis.
 ORGANISM Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Mycobacteriales; Corynebacteriineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 15916)
 Cole,S.T., Broesch,R., Parkhill,J., Garnier,T., Churcher,C.,
 Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
 Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
 Connor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S.,
 Hamlin,N., Holroyd,S., Hornsby,T., Jagals,K., Krogh,A., McLean,J.,
 Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
 Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
 Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
 Barrell,B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum: [[published erratum appears in Nature 1998 Nov
 12:396(6707):190]]
 2 (bases 1 to 15916)
 Parkhill,J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1460072.

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)
 CBS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

COMMENT

gene
IAEVLFLVIAINTLKRREAPAPAPPSAEVELLIEIRDLTKSCTRIIRPNC"
6236..6712
/gene="smpB"
CDS
6236..6712
/gene="smpB"
/note="small protein homolog"
/codon_start=1
/transl_table=11
/product="SmpB"
/protein_id="AA04322.1"
/db_xref="GI:6136301"
/translation="MADKNERPIKVMENKARPNVAIEDTIEAGIALTGTETKSTRN
GKSTIASYADSKNGETWLNATIPETLOGNRNHEKPRKLLHRRKRLIGVD
REGWTLPLKLTNGRRAKLQLAVALGKKLHDKRETKRDMWEPGRSRACGARG"
6848..7297
/gene="bcpB"
6848..7297
/gene="bcpB"
/note="bacterioferritin-comigrating protein"
/codon_start=1
/transl_table=11
/product="BcpB"
/protein_id="AA04323.1"
/db_xref="GI:6136302"
/translation="MLSALRGRTVFAVPTGEGPKISLVDDMDMIPGARGCTPQCA
FRDLFAELKAAGAHHVGLSTOSNEYOTEMASKLHPPIVLSPEKIALAKLPTME
VAGLMLKRLALTIIDAKITKTFYPVFPEDRNAGVDLWIKANPVKD"
complement(7306..8157)
/note="ORF283; contains two putative transmembrane
domains"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA04324.1"
/db_xref="GI:6136303"
/translation="MIADFCARPTSSPTSSSCNPSSAATRPNAARRAIVASAPAT
PRNSSARPRPMGAMTTLRSENAARGROPPTLVPRDCPLCRLVAPEGESRARAV
VGITPRSPEDVGRARLLVGLAPMGQANRGTGPTGDYAGDLVLTLEYEPAGVAG
QARPDDGLKIVDIRIANAVHCVPPONKRPAPETNCRFLANLIATMPLRLIIVAGR
IAHDTVLKPLNLKGSQAPFGHGAHVHAGAFRLYDSHCSRNTNCTVLPDPFRSVFS
KVKADLD"
complement(8154..8603)
/note="ORF149"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA04325.1"
/db_xref="GI:6136304"
/translation="MSPTNKIALFIDGANYAFAKTLGPDIDYKRLLKEFQSRGTL
LRAFTYALIEDQESSIRPLIDMDYNGYTVATKEFIDASGRKRYKGMDEILA
VAMELAHIDDMVLFSSGDGDFRSIVEAVQRRGVTVSRSPASRR"
8929..9243
/gene="rnpO"
8929..9243
/gene="rnpO"
/note="RNA-dependent nucleotide polymerase omega subunit
homolog"
/codon_start=1
/transl_table=11
/product="RnpO"
/protein_id="AA04326.1"
/db_xref="GI:6136305"
/translation="MARVTEPCIDKVNRPDLVLLAARAMISGSSQLVYDRNDK
NPVSLREIADTTISPEDLRELGALAEVVRGRRAAGYCAADRFRRRHRCGRYS
CR"
9484..11823
/gene="relA"
9484..11823
/gene="relA"
gene
9484..11823
/gene="relA"
CDS
9484..11823
/gene="relA"

Query Match 60.0%; Score 15; DB 1; Length 35209;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 cagtgatcgccag 21
Db 169 CAGTGATCGCCAG 183
RESULT 41
AC009603 35573 bp DNA INV 31-MAY-2000
LOCUS AC009603/c
DEFINITION Leishmania major chromosome 35 clone L3342 strain Friedlin,
complete sequence.
AC009603
VERSION AC009603.2 GI:6007874
KEYWORDS HTG.
SOURCE Leishmania major.
ORGANISM Leishmania major.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 35573)
REFERENCE
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
2 (bases 1 to 35573)
REFERENCE
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P. and Stuart,K.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
3 (bases 1 to 35573)
REFERENCE
AUTHORS Myler,P.J. and Sunkin,S.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2000) Seattle Biomedical Research Institute, 4
Nickerson Street, Seattle, WA 98109-1651, USA
4 (bases 1 to 35573)
REFERENCE
AUTHORS Myler,P.J.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4
Nickerson Street, Seattle, WA 98109-1651, USA
On Oct 4, 1999 this sequence version replaced gi:5801655.
FEATURES
source
1..35573
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:3664"
/chromosome="35"
/clone="L3342"
/gene="STPK5"
/note="L3342.2L; predicted using Glimmer, Testcode and
CodonUsage"
1..595
/gene="STPK5"
/note="Blastp similarity to serine-threonine protein
kinases from several organisms; contains Eukaryotic
protein kinase domain"
/codon_start=2
/product="Serine Threonine Protein Kinase 5"
/protein_id="AA070529.1"
/db_xref="GI:7882847"
/translation="IIALPPLPGPSNELDQVRIHNVIGTPPEILERLKRGSTHMDY
DEPKQGTGKGLPHYSARALDLMKKLLTYDEORCTAKEARLRAHYFSKLRPADKRS
HLIKHSISIRPTTVDDAASIGLGSSPRKTTNGPRLNSTITLASRKLPIVDKSPPT
KSCIVYNOSTLHGVHGAATVSASGDVLAKSLPLKL"
3863..4618
/gene="L3342.5"
/note="predicted using Glimmer, Testcode and CodonUsage;
CodonUsage not convincing at 5' end"
3863..4618
/gene="L3342.5"
/note="TBlastN similarity to several T. brucei GSS clones"
/codon_start=1

```

/product="L3342.5"
/protein_id="AAF70530.1"
/db_xref="GI:7882848"
/translation="MHCGPRGILGSGALTPYVADLTSCVSTIGRASPLGFSTP
PDKHPRIKQVGTGVAIQHVPGTTLKLTNRGRFTFSPVLAOLTHPPVAKATG
WHEIDTSDSDDDDDMPSEVDECLAAASVAGDEVELNOELRALQOAFYLCID
VLMDTSGMKTSMYSSEINNGPDYELDRKTRRRHMLAIRHREYDKELLMSDAA
DDGASASVETASVNSVNRQTSRCCRLRT"
5035..6909
/gene="L3342.3"
/note="predicted using Glimmer, Testcode and Codonusage;
ORF starts at 4615, Testcode & Codonusage predicts 3rd ATG
is used as start codon"
5035..6909
/gene="L3342.3"
/note="contains Glycican Pfam motif"
/codon_start=1
/product="L3342.3"
/protein_id="AAF70531.1"
/db_xref="GI:7882849"
/translation="MQICPEBGSDSASELLRESDDAGAVREVLDALCRVPASLF
SDALAEVDEAEVLCVLTATETQEDNAAFVOWGSEVMSRPERATTEGRIYLOH
FTRLVLCRRLHHDGPPPTLLSGDAENIPVLDQLYDHNHPEVRIYQIGRANA
QGLRYAGLQSTASTMAELLRLSEANOPMAAGARGLCDLRLHVAOKALGSKTL
EDVTRALPLLAOMGKTSEOOVKSTYERILTTAMSTTAVLAOOPKPACTIALIEGA
GCHYTPSSYKLEVMYLRMMRGAFETLSDAGRCILAMLYGVPKQVSOVKOATAMV
VAAYDASDSTADAVLAVLRALRYCTISFASIVGLYSKSTPLMONTANWOPSEHVA
VLAASAAAEETANSEIGIAVSTTARERLGLSPVHPSPVAPHSFACVGEADAYO
CVTCAALDVSPPALTMAMHADIDARVHSLCTPALAVSTMEALFAMSIAISRMT
AYVYLDITLQOMORPLEDDAETAVRCLEORAPAVTRIASVLEHEDYRLDQTL
TGAIVVDDAARAPVCTEQRSLEKQTSSEGTETRLRYCAALQHASREMLSTSAVY
"
7592..9346
/gene="L3342.4"
/note="predicted using Glimmer, Testcode and Codonusage,
but Testcode and Codonusage not convincing, especially at
5' end"
7592..9346
/gene="L3342.4"
/codon_start=1
/product="L3342.4"
/protein_id="AAF70532.1"
/db_xref="GI:7882850"
/translation="MLRCETEMTNAVESDVSALRHSSNRSNSTRATLRGAAPR
GSSKAYGMLEFSRROOLKSKGNTVASTGYAADAVIDOTRQDTPADRYTGFEL
DAPAMHAMATTKKQSNRPLRPGSTERTIVSRI SPRVALHERSDSTVYGDYLT
RLQALNACDTSOSLPSPEDQVWQHRPAMIAALPCRRPRGRSGSGWQSVRST
TPSLGVTTLVKEWKAWEKATAVLTGSTEVCMEKSLTLPFAFCMPSPRAGQSTL
SCVARLQSKRPVSAWMPDEPCASAGRDVAEEAOPWHRHLAEPLOSRLK
DAVSPVKPPPSILVPIGASIRIDNLERIRHFEHQRILKLOGELDKLEMRHRS
ALLRHNSKEDAAVYPAHSPSTASTPQAGERAMRPLLAEVFGPSGSAARHOLY
ESPILYSLMGSITWSEHARQAVOQESRVALLOSIVVTSAAVLTILAEMKRISRT
IRLWHPRENSGSHHALLHREOKFLKLMGRYEHTSGLLGAVDTVRPIDIDALPQO
AASFARAYAEVLTAAE"
10834..12516
/gene="PAB1"
/note="L3342.1: 511.1L; predicted using Glimmer, Testcode
and Codonusage; Glimmer predicts start codon at 10885"
10834..12516
/gene="PAB1"
/note="binds Poly(A) tail of mRNA; Blastp similarity to
many polyadenylate-binding proteins from several
organisms; contains RNA recognition motifs and bacterial
histidine kinase domain"
/codon_start=1
/product="PolyA Binding Protein 1"
/protein_id="AAF70533.1"
/db_xref="GI:7882851"
/translation="MAAAVQRAAPVAHOPOMDKPMOIASIYVGDIDATINEPOLVEL
FKPGTILINRYCKDITITQSLGTYGVNFDHDSAEKATIESMNRKGVGDKVRLMWOQ
RDPALRYSGNNGVFNKLEKVDYSKLHDIPTKGSILSCVWMDREKSGSGYVHF
KDEITSADKAIKMGADHASEDKKALVAVFIRRNALALVANFTNTVYIKOVLPTVY
NKDIEKFFPAKFGITSAAACKDKSGRVFAFCNFEKHDDAVKAVEAMHDDHIDITAP

```

```

GKLYVQAPQPSERERLIALROKYMHOALNNILYVRNEDPEFTGADLLELEKEYGEVK
SCRVAVSESGVSGFGEFSESNADENALRMNGMLNGKPLIYNIRORROVRYTML
RLQFOORLOMMROMHOPMPFVCSOCRRPGRGSGROOLCGRAGOHMPMPSPQOPAP
AOPQGRATPSAVGVOATKRHSBGDVPETPLPPIITPOLSEMSQOEORALADGDLFL
KVEILAEPLAKPTTGMLEMKPEAVEALLNDOKRLERVERTALCVLKAHQTA"
15582..15770
/gene="511.2"
/note="L3342.6: predicted using Glimmer, Testcode and
Codonusage; Glimmer predicts start codon at 15624"
15582..15770
/gene="511.2"
/product="511.2"
/codon_start=1
/protein_id="AAF70534.1"
/db_xref="GI:7882852"
/translation="MISNRYIAOHGSCMEVVPNGHEAEVDAALGFRPPNAPMHTHT
VFRPSKNLTATFTTAASSALGVHNTYQTLTPAALNLVNHGLRVADPESSRPHPOD
AKSATSLPLLPADRLSVGSDGLRGGTEVEREWKAPFTNSTSPNGAEGGDAA
AINGNHSLELVSTAPASARQESGOSTSALVPATLVPAKTAFASSDHCPSVSV
SPSLRLVRKSTSTASFPLOAPEVSAFTYVATVYLRDVLMOQRYLDKSOISGDA
RALPAPPAPEPPEAESQHOQOTAVLIATSTGSPANEDVMEDESDCYPLASTHAA
LVLSAERVRVGPVGAASPTFAAKVADVAAPDETRVNLAVSOPSAIMERRLR
GPATRAPMBGLTSPPAHSLGCAETGAMKTSRRIAEVPSAASVASTVQTLTAA
GQLORSSSPNAGGGEHPSLRAPAOESAGSVSRVHEPDLIPGLAEFLRAGNPSF
TSLSATHAAPPARAKASPTRRRRSGPPAADSPAVLIGHSPILORRGSGTA
LSQADAPVPSLHLPYAHREGSATVADVSPRGSARARVDSPTVAPQTLISLKS
EVSADDFPMSRVASPTPSLKDPVTASRDAASEPALL"
19700..21268
/gene="511.3"
/note="Orf1pha; L3342.7: predicted using Glimmer,
Testcode and Codonusage"
19700..21268
/gene="511.3"
/note="Prosite predicts bipartite nuclear localization
signal"
/codon_start=1
/product="511.3"
/protein_id="AAF70535.1"
/db_xref="GI:7882853"
/translation="MLLSDAEAKITVAVLSLDDPTTYEVENLYOLNCERTAAVRAA
LADMSGDYVQOLCEKGEAEKQVQARRAGVASTAADAAVSLDFALAEASQNAEI
ESAISAAMHGRPRKKSQGEKSKKEGGERHKKRRLDDEPLAEVPLPAPAGGSCEN
NAADKATSPSTRWVLSRDDLTSEVPLALSTSDRVCAEVPNTPPLARFTEBEVT
RDSASOCAYELTACVLIAMRYGEPOLSNABCEKMMQSOVAASAKGTIEDGPPRS
AKAARASRPTSSDQGBGADSTSLVEMTSLBOLIAQYOAOAEEBEDDADTLT
LPRLISERFOQRHGFTEDEWVTLTCHGGTFAGGVAKSTCTHKAFORTVVRKQGG
KQSSNKKDAGSYNSVGSQIRAVQDEVKRVADVRDILLMPYITQASFTILYAAPQONR
AVLTDSLPLPAVASVSGRGKGVSPITQKDPVRSVPLTTRPTFEVRYRIVGCSRSL
LYVROEAOORNTN"
23224..23730
/gene="511.4"
/note="ORF2; L3342.8: predicted using Glimmer, Testcode
and Codonusage"
23224..23730
/gene="511.4"
/codon_start=1
/product="511.4"
/protein_id="AAF70536.1"
/db_xref="GI:7882854"
/translation="MSDYSDDDDVVVLQVCANRRCOGIDTLEFDEETNOSYCGRCA
LYARTEGFRVLISNDIDGLVYKILIDQDENSGRWTKENNAFQEVTERSDPTE
SSAALKEFFEEEDIDISAHGKGAVALLVLDENMYGVLNNVVALVDESEAMEYQ

```

Query Match 60.0%; Score 15; DB 4; Length 35573;
 Best Local Similarity 100.0%; Pred. No. 2; 7e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ccgacagatgcatc 16
 Db 7912 CGCAGCAGTGCATC 7898

RESULT 42


```

/protein_id="AA69577.1"
/db_xref="GI:7769949"
/translation="MSGSGATGSGSAGLSSSTNTNCKTSTASRSGNSTNNISITG
KKRSTTAATAATTPEACGSEGCYHSSRSTSPPTSDDEVASLTAAATAGSRKSLPT
PRTGATSCESAMTLHSPAGTPARAKRPPPEATSLLEKVLTAQALGARNMEQ
GDRVVESTPTIIVIAAGKADTQREERTAEITSSGADAAATASGNDGAVAKRKY
ATSDASALNTVSEFEEDTADLEITLEDITISNDPPEGCCIVNRSFKQSLCTIM
YASEGYSASMSALVALOOKSYPRFEDLERRRPRSEKTAALPGDSTSLPSTKAT
DKTEDVLPRTSYTHRTKYPPLGFIHMLSLSTIEHRATNPALYADIEVLKMLM
ANADLDYLSLDETPPLLSGLQCTAAPSWSVPEPLRLHLIPEEVTITRDRWGF
LQNTVLPSCGAOSVALLAKYATMSRVRNDIVCALICDSEFSEACFALVLPVPT
PAMASGSSARNRGLSVEASLTSLTSGSCSAGSGVNGAGSPSTPAPVTPVPAQ
KAELOVLAHLREFRELVTLSIAELGRENGLPYGKTYGSEPLISLIVERAMPMT
AAAFVHMDGSGGALNGARNGGSSANSNAAGSTSTANCYSLAVEDEIAVL
DVTVLRNROEQLMNEHVRNLIADPKKXNYTFPMRLIOLVGPICATAAATSGTG
TNATRSLSITPPPKLIDRSMSTNPFALFVGLHDEGSAASERKLDSESTGRN
REFNLTITKTYIAHATRGVMPVPLPSLGGVSTENIPIVOFNANNAQNGARVH
ODNGSDTASATVGAENTCPAAGGGSGAASVPLPGITPPLALIVLEMYNLTNA
NRELILKTVHYSKSHIMHIEGVCDTLCENAGARCOIGLVLGCVEFLVYNTOM
TITTAEDENPLSVASRREPLSPOLVTTICROTIVERTDPTGFLKAYNELGRRDS
VFHSLSLVDITIGKSVHPEDSTNNLRDVRDYLFFHFRLEPREVRCRREALS
EYTVRIGORATSPNDVSYSVLSSTAGSELTRKCSKURFVDEDAVSGDGVVSV
VSPALVWMAAEEKORRRHRSMEQMNATANIANGSPLQVRSASGLADHDGTRA
SPSSSPSRKSSPLTISITRPGSSNSGSGASSTPAGDILKVAAGHNSSYNGD
GPLAQSPPSTVPTTRKADYLRPTAPKEKPPAPRFAVLENSPSTNAASGEO
GKRSRMPNATAKHAPPEKVLSTRTGGSSVQSEVHANGTALPESDNYVLPKIKKL
LSTGCKRS"
complement(32397..35102)
/gene="LA286.3"
/feature="predicted using Glimmer, Testcode and Codonusage"
complement(32397..35102)
/gene="LA286.3"
/codon_start=1
/product="LA286.3"
/protein_id="AA69576.1"
/db_xref="GI:7769948"
/translation="MASASSAAVRAAARDNPSPPPPRTAORPRLVGLQCTP
SRSLKFSASMAHROFLADPTAISRDELFLFMLEFVPRQOIVDRVLPDPATYAP
ELLEERVITDPRHAYGQAGLPAKSTIPSTNIVSYSGTIEVPATSCSRPTTMGCP
AYGDVYALDAFEVNGRGRANDPRGVGTIOANLSABNRNNNGESTALVRSQINAG
BELMKSAGHRLSATWMTVOGEPTMRRLGCVLPESFRETYSADSDTSASAPAL
ATSLPQKASDASNAEDSCAAGSGLTASVTDGASVSKVDGSGSADPIDAGTLLS
ADTKOPLARSLRVLPGANDLMECTOCGMSICERSTADALRCACATPLGSLARL
VELLSPATTOAMLALAEAGLDAGOCMASTSTPOTOVASVITSTSTSSRRTMPM
NVPLPMQVWDAVPLVTLAKHSREFTOHITFYVYDGLDREKGRSTRDHGIGPT
VAAPMPGTPDAEQSGEACPSKSGSRQPHSSADASLAWVRDERPLVTEENAGR
GRGSAHSGSGSRNALMERQDLDSELAITLPGIALPLLSVSVISPAHALAS
AASPTKLEPLLEWTASAGRGTSAGHSGVQGTSTALVSDTSDSVSITRPLS
YLSLTQVEGRLLVLRRAAGSPSPAAGTYPSPOMTIVRRLOCMTRLETGAKROP
DVASYGVGAVREGDTRCPVNSCLEILPLRFLPRLRHRCGRGACIARSGSLCDA
EBAALOTPCDRLDRLSLVYTNFMYCPCPLVLEVTWKKNSVNAAGTVEVSCADSVGE
DEVDDALASCNVALVITMDALCSFACAVATIPICAFELAKAQ"
complement(36081..36680)
/gene="LA286.2"
/feature="predicted using Glimmer, Testcode and Codonusage"

```

Query Match 60.0%; Score 15; DB 4; Length 41193;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3 gcaagcagtgacatcg 17
|||||
Db 9371 GCAGCAGGTGCATCG 9385

RESULT 43
LOCUS AC010753 44545 bp DNA HTG 18-FEB-2000
DEFINITION Drosophila melanogaster clone RPI198-10P16, *** SEQUENCING IN
ACCESSION AC010753
VERSION AC010753.2 GI:6996807
KEYWORDS HTG; HTGS_PHASE1.

```

SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 AUTHORS
 Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
 Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunc, C.,
 Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
 David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Pashid, N.,
 Dugan-Hoch, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
 Forum-Tansey, J., Frantz, P., Ganseth, R., Gorrell, J.H., Gorrell, L.L.,
 Guenard, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,
 Holloway, C., Hoesak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
 Kelly, S., Kondelowski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
 Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
 Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
 Owsal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
 Quides, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shan, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,
 Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahban, A.,
 Watlington, S., Weinstock, G., Weinstock, I.R., Williamson, A.,
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 44545)
 REFERENCE
 AUTHORS
 Worley, K.C.
 TITLE
 JOURNAL
 COMMENT
 Submitted (22-SEP-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 18, 2000 this sequence version replaced gi:5917930.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: DRHM
 Center clone name: RPI198-10P16
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 99% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 10931 bases at least Q40
 Consensus quality: 20681 bases at least Q30
 Consensus quality: 27126 bases at least Q20
 Estimated insert size: 36762; sum-of-contigs estimation
 Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1244: contig of 1244 bp in length
 * 1245 1264: gap of unknown length
 * 1265 1280: contig of 1316 bp in length
 * 1281 2581: gap of unknown length
 * 2582 2600: gap of unknown length
 * 2601 3779: contig of 1179 bp in length
 * 3780 3799: gap of unknown length
 * 3800 5213: contig of 1413 bp in length
 * 5213 5233: gap of unknown length
 * 5233 6780: contig of 1548 bp in length
 * 6781 6800: gap of unknown length
 * 6801 8288: contig of 1488 bp in length
 * 8289 8309 9348: gap of unknown length
 * 9348: contig of 1040 bp in length


```

* 9349 9368: gap of unknown length
* 9369 10616: contig of 1248 bp in length
* 10617 10636: gap of unknown length
* 10637 11240: contig of 604 bp in length
* 11241 11260: gap of unknown length
* 11261 12652: contig of 1392 bp in length
* 12653 12672: gap of unknown length
* 12673 14216: contig of 1544 bp in length
* 14217 14236: gap of unknown length
* 14237 15757: contig of 1521 bp in length
* 15758 15777: gap of unknown length
* 15778 17262: contig of 1485 bp in length
* 17263 17282: gap of unknown length
* 17283 18374: contig of 1092 bp in length
* 18375 18394: gap of unknown length
* 18395 19394: contig of 1000 bp in length
* 19395 19414: gap of unknown length
* 19415 20608: contig of 1194 bp in length
* 20609 20628: gap of unknown length
* 20629 21616: contig of 988 bp in length
* 21617 21636: gap of unknown length
* 21637 22529: contig of 893 bp in length
* 22530 22549: gap of unknown length
* 22550 23984: contig of 1435 bp in length
* 23985 24004: gap of unknown length
* 24005 24950: contig of 946 bp in length
* 24951 24970: gap of unknown length
* 24971 25858: contig of 888 bp in length
* 25859 25878: gap of unknown length
* 25879 27488: contig of 1610 bp in length
* 27489 27508: gap of unknown length
* 27509 29103: contig of 1595 bp in length
* 29104 29123: gap of unknown length
* 29124 30355: contig of 1232 bp in length
* 30356 30375: gap of unknown length
* 30376 31235: contig of 860 bp in length
* 31236 31255: gap of unknown length
* 31256 32692: contig of 1437 bp in length
* 32693 32712: gap of unknown length
* 32713 33672: contig of 960 bp in length
* 33673 33692: gap of unknown length
* 33693 35335: contig of 1643 bp in length
* 35336 35355: gap of unknown length
* 35356 36573: contig of 1218 bp in length
* 36574 36593: gap of unknown length
* 36594 38685: contig of 2092 bp in length
* 38686 38705: gap of unknown length
* 38706 41189: contig of 2484 bp in length
* 41190 41209: gap of unknown length
* 41210 44545: contig of 3336 bp in length.

```

```

FEATURES
  source
    1..44545
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="RPC198-10F16"
BASE COUNT 11895 a 9936 c 9918 g 12034 t 702 others
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 61; Length 44545;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 cagcaggtgcatgc 18
  |||
Db 9470 CAGCAGTGCATCGC 9484

```

```

RESULT 44
AC014684/c AC014684 55307 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.

```

```

ACCESSION AC014684
VERSION AC014684.1 GI:6436651
KEYWORDS HTG; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 55307)
REFERENCE Adams,M. and Venter,J.C.
  Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
  Rockville, MD, USA
  Direct Submission
  TITLE
  JOURNAL
COMMENT
  This sequence was identified as CDM:10210542 by the submitter.
  For further information on this sequence e-mail to fly@celera.com.
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
  Location/Qualifiers
    1..55307
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT 16942 a 11043 c 11010 g 16312 t
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 63; Length 55307;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 agcaggtgcatgcc 19
  |||
Db 6293 AGCAGTGCATGCC 6279

```

```

RESULT 45
AC014903 AC014903 73691 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC014903
VERSION AC014903.1 GI:6436432
KEYWORDS HTG; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 73691)
REFERENCE Adams,M. and Venter,J.C.
  Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
  Rockville, MD, USA
  Direct Submission
  TITLE
  JOURNAL
COMMENT
  This sequence was identified as CDM:10210292 by the submitter.
  For further information on this sequence e-mail to fly@celera.com.
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
  Location/Qualifiers
    1..73691
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT 20452 a 16659 c 16681 g 19899 t
ORIGIN

```

```

FEATURES
  source
    1..73691
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT 20452 a 16659 c 16681 g 19899 t
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 63; Length 73691;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 cgcagcaggtgcatc 16

```

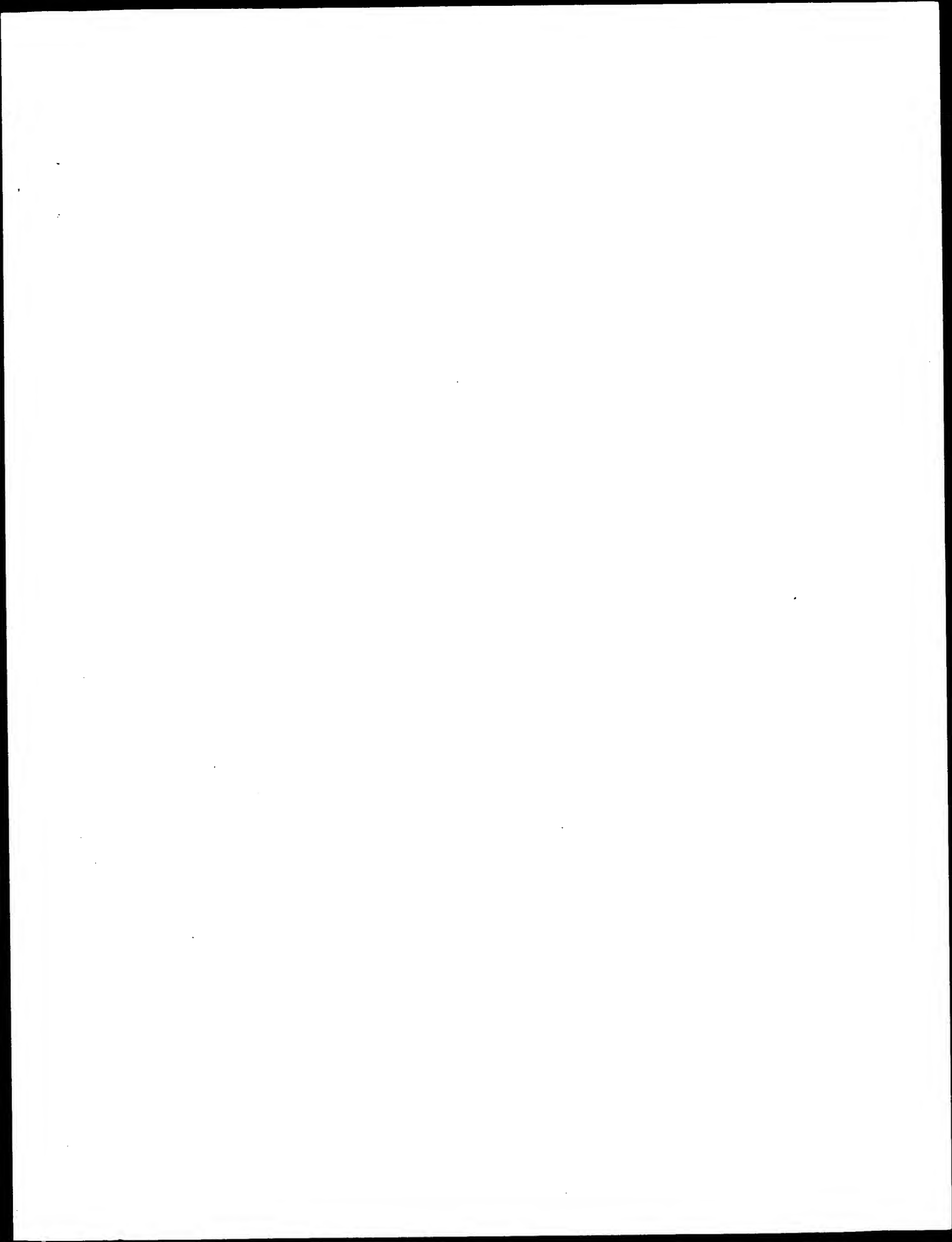

Wed Oct 10 07:45:52 2001

us-09-396-196f-5.oli.rge

Page 41

|||||||
Db 24505 CGACGAGGTGCATC 24519

Search completed: October 9, 2001, 15:45:48
Job time: 14734 sec



XX Example 2: Column 37-40; 34pp; English.
 PS
 XX

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC

XX
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgacgaggtgcatcgccagcatt 25
 ||||||||||||||||||
 Db 79 gccgacgaggtgcatcgccagcatt 103

RESULT 2

AA091329
 ID AA091329 standard; DNA; 1084 BP.

XX
 AC AA091329;

XX 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

PH Key Location/Qualifiers
 FT CDS 24..1064
 FT /tag=a

PN GB2216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

XX The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgacgaggtgcatcgccagcatt 25
 ||||||||||||||||||
 Db 102 gccgacgaggtgcatcgccagcatt 126

RESULT 3

AA060496
 ID AA060496 standard; DNA; 1121 BP.

XX
 AC AA060496;

XX 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetase enzyme; E.coli; desthiobiotin; ds.

PH Key Location/Qualifiers
 FT CDS 42..1082
 FT /tag=a

PN JP61149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

XX WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

XX The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.

SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.00017;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgacgaggtgcatcgccagcatt 25
 ||||||||||||||||||
 Db 120 gccgacgaggtgcatcgccagcatt 144

RESULT 4

AA062386
 ID AA062386 standard; DNA; 5872 BP.

XX
 AC AA062386;

XX 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

KW promoter plac; biotin synthase; KAPA synthase;
 KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126265.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128233.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139452.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140333.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.

PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 23-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0148723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151338.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156658.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcacgcacgca 23
|||||||
Db 84 aggtgcacgcacgca 99

RESULT 6

AAC36258
ID AAC36258 standard; DNA: 1500 BP.

XX AAC36258;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142194.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

```

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match Best Local Similarity 64.0%; Score 16; DB 21; Length 1500;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 8 aggtgcatgcgcagca 23
Db 84 aggtgcatgcgcagca 99

```

RESULT 7

ID AAF11615 standard; cDNA: 612 BP.

XX AAF11615;

DT 13-MAR-2001 (first entry)

DE Aspergillus niger EST SEQ ID NO:4138.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

OS metabolic pathway engineering; catalytic pathway engineering; ss.

PN Aspergillus niger.

PD WO200056762-A2.

XX 28-SEP-2000.

PR 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 XX
 DR
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS
 XX
 PS Claim 87; Page 1813; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Pisarium venetum*; AAF11248 to AAF1853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 612 BP; 147 A; 162 C; 144 G; 156 T; 3 other;

Query Match 60.0%; Score 15; DB 21; Length 612;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcaggtcatcg 17
 |||||
 DB 118 GCACGACGTGCATCG 104

RESULT 8
 AAC93313
 ID AAC93313 standard; cDNA; 2597 BP.
 XX
 AC AAC93313;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA sequence #4.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058495-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000WO-US07661.
 XX
 PR 26-MAR-1999; 99US-0126504.
 PR 07-JAN-2000; 2000US-0174847.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI; 2000-611720/58.
 DR P-PSDB; AAB51383.
 XX
 PT New nucleic acid molecules encoding 45 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS
 XX
 PS Claim 1; Pages 327-328; 410pp; English.
 XX
 CC The invention relate to the isolation of genes AAC93310-C93354 encoding
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 SQ Sequence 2597 BP; 721 A; 604 C; 534 G; 737 T; 1 other;

Query Match 60.0%; Score 15; DB 21; Length 2597;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagcaggtcatcgcca 20
 |||||
 DB 2256 gcagcaggtcatcgcca 2270

RESULT 9
 AAA01491/c
 ID AAA01491 standard; cDNA; 300 BP.
 XX
 AC AAA01491;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
 XX
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO958675-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 13-MAY-1999; 99WO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085337.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kila D, Garcia V, Jones LM, Stache-Crain B;
 XX
 DR WPI: 2000-126369/11.
 XX
 PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 XX
 PS Claim 1; Page 536; 1097Pp; English.
 XX
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SQ Sequence 300 BP; 56 A; 89 C; 94 G; 61 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatc 16
 |||||
 DB 108 GCAGCAGGTGCATC 95

RESULT 10
 AAF71326
 ID AAF71326 standard; DNA; 667 BP.
 XX
 AC AAF71326;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:377.
 XX
 KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
 KW Brevibacterium; environmental condition; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100842-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000MO-IB00911.
 XX

PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032127.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032129.
 PR 09-JUL-1999; 99DE-1032226.
 PR 14-JUL-1999; 99DE-1032920.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1032930.
 PR 14-JUL-1999; 99DE-1032933.
 PR 14-JUL-1999; 99DE-1032935.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033002.
 PR 14-JUL-1999; 99DE-1033003.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041390.
 PR 31-AUG-1999; 99DE-1041391.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroege B, Schroeder H, Zelder O, Haberhauer G;
 PI WPI: 2001-061974/07.
 PT P-PSDB; AAB79211.
 PT
 PT New isolated Corynebacterium glutamicum nucleic acid for production or
 PT modulation of production of fine chemicals such as amino acids,
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
 PT or enzymes -
 XX
 PS Claim 3; Page 630-631; 712Pp; English.
 XX
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host
 CC cells and production of fine chemicals, such as, an organic acid,
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
 CC be modulated. The presence of (I) or HA proteins encoded by then are
 CC used for diagnosing the presence or activity of Corynebacterium
 CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
 CC used as markers for genetically engineered Corynebacterium or
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
 CC homeostasis in C. glutamicum or help the microorganism to adapt to
 CC different environmental conditions.
 XX
 SQ Sequence 667 BP; 146 A; 210 C; 172 G; 139 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 667;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgca 14
 |||||
 DB 168 gcgcagcaggtgca 181

RESULT 11
 AAV20504/c

```

ID  AAV20504 standard; DNA; 719 BP.
XX
XX  AAV20504;
XX
XX  30-JUN-1998 (first entry)
XX
XX  Conus geographus conantokin DNA.
DE
XX  Conantokin; predatory cone snail; treatment; neurologic disorder;
XX  psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
XX  HIV infection; ophthalmic indication; memory; learning defect;
XX  cognitive defect; ss.
XX
XX  Conus geographus.
OS
XX
XX  Key      Location/Qualifiers
XX  CDS      110..412
XX           /*tag= a
XX           /product= conantokin
XX
XX  MO9803541-A1.
XX
XX  29-JAN-1998.
XX
XX  21-JUL-1997; 97WO-US12618.
XX
XX  22-JUL-1996; 96US-0684742.
XX
XX  (COGN-) COGNETIX INC.
XX  (UTAH ) UNIV UTAH RES FOUND.
XX
XX  Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;
XX  Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;
XX  Walker C, Zhou L;
XX
XX  WPI; 1998-120694/11.
XX  P-PSDB; AAM48210.
XX
XX  New conantokin peptide(s) - useful for e.g. treating neurologic or
XX  psychiatric disorders, or the management of pain
XX
XX  Claim 19; Pages 79-80; 122pp; English.
XX
XX  The present sequence encodes Conus geographus conantokin, peptide
XX  derivatives of which can be used to treat neurologic and
XX  psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
XX  or analgesic agent. Neurologic and psychiatric disorders include
XX  epilepsy, convulsions, neurotoxic injury (associated with
XX  conditions of hypoxia, anoxia or ischaemia, which typically follow
XX  stroke, cerebrovascular accident, brain or spinal cord trauma,
XX  myocardial infarct, physical trauma, drowning, suffocation,
XX  perinatal asphyxia or hypoglycaemic events), neurodegeneration
XX  (associated with Alzheimer's disease, senile dementia, Amyotrophic
XX  lateral sclerosis, Multiple Sclerosis, Parkinson's disease,
XX  Huntington's disease, Down's Syndrome, Korsakoff's disease,
XX  schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
XX  dementia and neuronal damage associated with uncontrolled
XX  seizures), chemical toxicity (such as addiction, and morphine,
XX  opiate, opioid and barbiturate tolerance), pain (acute, chronic,
XX  migraine), anxiety, major depression, manic-depressive illness,
XX  obsessive-compulsive disorder, schizophrenia and mood disorders
XX  (such as bipolar disorder, unipolar depression, dysthymia and
XX  seasonal affective disorder) and dystonia (movement disorder),
XX  sleep disorder, muscle relaxation and urinary incontinence. The
XX  peptide can also be used to treat HIV infection, ophthalmic
XX  indication and memory, learning or cognitive defects.
XX
XX  Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;

```

```

OY  7 caggtgcatcgcca 20
    |||
DB  118 CAGGTGTCATCGCCA 105

RESULT 12
AAV17130/C
ID  AAV17130 standard; DNA; 719 BP.
XX
XX  AAV17130;
XX
XX  30-JUN-1998 (first entry)
XX
XX  Conus geographus conantokin DNA.
DE
XX
XX  Conantokin; predatory cone snail; treatment; neurologic disorder;
XX  psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
XX  HIV infection; ophthalmic indication; memory; learning defect;
XX  cognitive defect; ss.
XX
XX  Conus geographus.
OS
XX
XX  Key      Location/Qualifiers
XX  CDS      110..412
XX           /*tag= a
XX           /product= conantokin
XX
XX  MO9803189-A1.
XX
XX  29-JAN-1998.
XX
XX  21-JUL-1997; 97WO-US12652.
XX
XX  06-DEC-1996; 96US-0762377.
XX  22-JUL-1996; 96US-0684750.
XX
XX  (COGN-) COGNETIX INC.
XX
XX  Layer RT, McCabe RT, Zhou L;
XX
XX  WPI; 1998-120469/11.
XX  P-PSDB; AAM49989.
XX
XX  Use of conantokin peptide(s) - for treating disorders involving
XX  excessive excitation of nerve cells by excitatory amino acids or
XX  agonists of the N-methyl-D-aspartate receptor
XX
XX  Example 4; Pages 79-80; 122pp; English.
XX
XX  The present sequence encodes Conus geographus conantokin, peptide
XX  derivatives of which can be used to treat neurologic and
XX  psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
XX  or analgesic agent. Neurologic and psychiatric disorders include
XX  epilepsy, convulsions, neurotoxic injury (associated with
XX  conditions of hypoxia, anoxia or ischaemia, which typically follow
XX  stroke, cerebrovascular accident, brain or spinal cord trauma,
XX  myocardial infarct, physical trauma, drowning, suffocation,
XX  perinatal asphyxia or hypoglycaemic events), neurodegeneration
XX  (associated with Alzheimer's disease, senile dementia, Amyotrophic
XX  lateral sclerosis, Multiple Sclerosis, Parkinson's disease,
XX  Huntington's disease, Down's Syndrome, Korsakoff's disease,
XX  schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
XX  dementia and neuronal damage associated with uncontrolled
XX  seizures), chemical toxicity (such as addiction, and morphine,
XX  opiate, opioid and barbiturate tolerance), pain (acute, chronic,
XX  migraine), anxiety, major depression, manic-depressive illness,
XX  obsessive-compulsive disorder, schizophrenia and mood disorders
XX  (such as bipolar disorder, unipolar depression, dysthymia and
XX  seasonal affective disorder) and dystonia (movement disorder),
XX  sleep disorder, muscle relaxation and urinary incontinence. The
XX  peptide can also be used to treat HIV infection, ophthalmic
XX  indication and memory, learning or cognitive defects.

```

XX Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 719;
Best Local Similarity 100.0%; Pred. NO. 80;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 cagtgcatcgcca 20
| | | | | | | | | |
DB 118 CAGTGCATCGCCA 105

RESULT 13

AAC98831 standard; cDNA; 1190 BP.

AAC98831;

09-MAR-2001 (first entry)

Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.

Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
detection; diagnosis; identification; cytostatic; neuroprotective;
nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
antiinflammatory; cardiant; gene therapy; chromosome mapping;
linkage analysis; tissue identification; tissue typing; forensic;
neural; immune system; muscular; reproductive; gastrointestinal;
pulmonary; cardiovascular; renal; proliferative; ss.

Homo sapiens.

WO200055320-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05989.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-579444/54.

P-PSDB; AAB54066.

New nucleic acid that is a pancreatic cancer antigen for preventing,
treating, or ameliorating a medical condition, particular pancreatic
cancer, or for use in assays for diagnosing a pathological condition -
Claim 1; Page 537; 1379pp; English.

AAC98773 to AAC99231 encode the human pancreatic cancer associated
proteins, called pancreatic cancer antigens, given in AAB54008 to
AAB54466. The human pancreatic cancer antigens have cystostatic,
neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
gynaecological, cardiant and antiinflammatory activities, and can be used
in gene therapy. The polynucleotide and proteins can be used for
preventing, treating, or ameliorating a medical condition or in assays
for diagnosing a pathological condition or a susceptibility to one in a
subject. Binding partners to the proteins and the activity of the
proteins can be identified. The pancreatic cancer antigens can be used to
detect, treat or prevent pancreatic disorders, especially cancer.
Agonists and antagonists to the antigens can be screened for. The
pancreatic cancer antigen polynucleotides can be used to design nucleic
acid hybridisation probes that can be used in chromosome mapping, linkage
analysis, tissue identification and/or typing and a variety of forensic
and diagnostic methods. The proteins can be used to generate antibodies
which are used to purify, detect and target the polypeptides, including
both in vivo and in vitro diagnostic and therapeutic methods. The
proteins can be used to treat or prevent neural, immune system, muscular,

reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
sequences used in the exemplification of the present invention.

Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1190;
Best Local Similarity 100.0%; Pred. NO. 79;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatcgccagc 22
| | | | | | | | | |
DB 1141 ggtgcatcgccagc 1154

RESULT 14

AAZ56381/C standard; DNA; 1344 BP.

AAZ56381;

17-MAR-2000 (first entry)

Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.
Flagellin; flc; antigen; detection; ds.
Escherichia coli.

WO961458-A1.

02-DEC-1999.

21-MAY-1999; 99WO-AU00385.

21-MAY-1998; 98AU-0003634.

(UNSY) UNIV SYDNEY.

Reeves PR, Wang L;

WPI: 2000-072598/06.

Novel nucleic acid molecule useful for the detection of flagellated
bacterial strains in food, faeces, etc.

Claim 3; Page 225; 245pp; English.

AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all
or part of an Escherichia coli flagellin protein except a protein
expressed by E. coli H1, H7, H12 or H48 type strains. The present
invention also describes a method of detecting the presence of E. coli
of a particular H serotype in a sample, comprising specifically
hybridising a nucleic acid, preferably at least a pair, derived from a
flagellating gene, specific for a particular flagellin gene associated
with the H serotype, to any E. coli in the sample which contain the gene,
and detecting any hybridised molecules, identifying the presence of that
serotype in the sample. (I) are useful for: (1) detecting the presence
of E. coli of H serotype in a sample by hybridising at least one or a
pair of (I) to any E. coli in the sample and detecting the hybridised
nucleic acid molecules; and (2) for detecting the presence of both O
and H-serotypes of E. coli by hybridising at least one or a pair of (I)
to any E. coli present in the sample and detecting the hybridised
nucleic acid molecules. (I) is particularly useful for detecting the
combination of O and H antigen. Hybridised (I) when using at least one
(I) is detected by southern blot analysis and, when using a pair of (I),
is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
represent primers used in the exemplification of the present invention.

Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gcatcgccagcatt 25
 |||||
 DB 590 GCATCGCCAGCATT 577

RESULT 15
 AAA69762/c

ID AAA69762 standard; cDNA; 2017 BP.

XX AAA69762;

AC AAA69762;

DT 07-NOV-2000 (first entry)

XX Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

DE Human; ovarian carcinoma; ovarian cancer; therapy; diagnosis;
 KW tumour antigen; identification; cytosolic; gene therapy; vaccine; ss.

XX Homo sapiens.

OS WO200036107-A2.

XX 22-JUN-2000.

PD 17-DEC-1999; 99WO-US30270.

XX 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404679.

XX (CORI-) CORIXA CORP.

PA Mitcham JL, King GE, Algate PA, Fridakis TN;

PI WPI; 2000-431589/37.

XX Immunogenic portion of an ovarian carcinoma protein and the nucleic
 PT acid encoding it, useful for the diagnosis, prevention and treatment of
 cancer, preferably ovarian cancer -

XX Claim 1; Fig 2; 299pp; English.

PS The present invention describes an isolated polypeptide comprising an
 CC immunogenic portion of an ovarian carcinoma protein (or its variants).

CC Ovarian carcinoma proteins, and polynucleotides encoding them, have
 CC cytosolic activity and can be used in gene therapy and vaccines.

CC Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines
 CC are useful for the prevention, diagnosis and treatment of cancer.

CC preferred ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557
 CC represent human ovarian carcinoma polynucleotides and proteins used in
 CC the exemplification of the present invention.

XX Sequence 2017 BP; 673 A; 440 C; 462 G; 442 T; 0 other;

QY Query Match 56.0%; Score 14; DB 21; Length 2017;

Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcatcgccag 21
 |||||

DB 964 AGGTGTCATCGCCAG 951

RESULT 16
 AAT30684/c

ID AAT30684 standard; DNA; 2193 BP.

XX

AC AAT30684;
 XX 11-SEP-1996 (first entry)

DE Kaposi's sarcoma associated herpesvirus ORF22.

KW Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;
 therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

XX Kaposi's sarcoma associated herpesvirus.

OS WO9615779-A1.

PN 30-MAY-1996.

PD 21-NOV-1995; 95WO-US15136.

PF 11-APR-1995; 95US-0420235.

PR 21-NOV-1994; 94US-0343101.

XX (UYCO) UNIV COLUMBIA NEW YORK.

PA Chang Y, Moore PS;

PI WPI; 1996-268320/27.

DR P-PSDB; AAR97832.

XX Herpes virus associated with Kaposi's sarcoma - also definitive DNA
 PT sequences, useful for diagnosis of and to develop prods. for
 PT treatment of Kaposi's sarcoma

PS Claim 17; Page 198-201; 277pp; English.

XX Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),
 CC 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2033-
 CC 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment
 CC of a newly identified human gamma-2 herpesvirus associated with
 CC Kaposi's sarcoma (KS). The ORFs were identified by sequence and
 CC positional homology to known herpesvirus sequences. Probes that bind
 CC specifically to the KS associated herpesvirus can be used for KS
 CC diagnosis. Antisense or triplex oligonucleotides are useful for
 CC prophylaxis or treatment of KS, and the protein products (see also
 CC AAR97830-46) of the 17 open reading frames are useful as vaccines.

XX Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

QY Query Match 56.0%; Score 14; DB 17; Length 2193;

Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcattcgccagcatt 24
 |||||

DB 46 TGCATTCGCCAGCAT 33

RESULT 17
 AAT16814/c

ID AAT16814 standard; DNA; 2193 BP.

XX AAT16814;

AC AAT16814;

DT 13-AUG-1996 (first entry)

XX Kaposi's sarcoma associated herpesvirus ORF22.

DE Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
 KW glycoprotein H; ss.

XX Kaposi's sarcoma associated herpesvirus.

OS WO9606159-A1.

PN

XX

XX

XX

XX

```

PD 29-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10194.
XX
XX 11-APR-1995; 95US-0420235.
PR 18-AUG-1994; 94US-0292365.
PR 21-NOV-1994; 94US-0343100.
XX
PA (UNCO ) UNIV COLUMBIA NEW YORK.
PA (GRAN/) GRANT D E.
PA (VIEL/) VIELE L.
XX
PI Chang Y, Moore PS, Grant DE, Viele L;
DR WPI: 1996-151362/15.
XX P-PSDB; AAR93608.
XX
PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated
PT vectors and proteins, used in detection and vaccination.
XX
PS Claim 17; Page 211-214; 305pp; English.
XX
CC Kaposi's sarcoma associated herpes virus (KHSV) clone K35 (AAT16806),
CC obtd. from a KS lesion genomic library, includes 15 complete ORFs and
CC 2 incomplete ORFs (AAT16807-23) named according to their herpesvirus
CC salmairi positional homologues. ORF23 (AAT16814) codes for
CC glycoprotein H (AAR93608). KSHV DNA may be incorporated into a vector
CC and expressed in host cells to produce peptides useful in vaccines or
CC for raising antibodies. The DNA may itself be used to immunise a
CC subject, or used to design therapeutic antisense and triplex
CC molecules or diagnostic probes, or to raise transgenic animals.
XX
SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
   |||||
Db 46 TGCATGCCACGAT 33

RESULT 18
AAZ20914/C
ID AAZ20914 standard; DNA: 2193 BP.
XX
AC AAZ20914;
XX
DT 02-DEC-1999 (first entry)
XX
DE Nucleotide sequence of gH glycoprotein of HHV-8.
XX
KW vaccine; antibody; Kaposi's sarcoma; glycoprotein; detection;
KW HHV-8; Human Herpesvirus type 8; ds.
XX
OS Human Herpes Virus type 8.
XX
PN WO9945948-A1.
XX
PD 16-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05464.
XX
PR 12-MAR-1998; 98US-0077670.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Whitbeck JC, Eisenberg RJ, Cohen GH;
DR WPI: 1999-551211/46.
DR P-PSDB; AAY42395.

```

```

XX
XX New nucleic acid molecules encoding human herpesvirus type glycoprotein
PT H and glycoprotein L polypeptides for diagnosis and treatment of
PT Kaposi's Sarcoma
XX
PS Claim 3; Fig 1; 75pp; English.
XX
CC This is the nucleotide sequence of the human herpes virus type 8
CC glycoprotein H. This sequence and it's encoded protein (AAY42395) can be
CC used in the development of vaccines, antibodies and other preparations.
CC The vaccine is useful as a therapeutic agent for prevention of HHV-8
CC disease and as a therapeutic agent for treatment of HHV-8 infection
CC in order to boost the immune response in the animal. Administration of
CC a pure preparation of soluble HHV-8 gH/gL complex or at least one
CC polynucleotide encoding the soluble complex is useful for preventing or
CC treating an HHV-8 infection in a human.
CC Diagnosing of an HHV-8 infection in a human is carried out by contacting
CC a biological sample with an antibody which specifically binds to either
CC glycoprotein or in complex form. If binding of the antibody occurs then
CC the human has HHV-8 infection.
CC Alternatively the infection is diagnosed by addition of one of the
CC HHV-8 glycoproteins or complex and determining whether specific binding
CC of the protein to an antibody in the biological sample occurs. A further
CC diagnostic method involves contacting the biological sample with an
CC HHV-8 glycoprotein H polynucleotide or HHV-8 glycoprotein L
CC polynucleotide and determining whether binding to a nucleic acid in the
CC sample occurs.
XX
SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2193;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
   |||||
Db 46 TGCATGCCACGAT 33

RESULT 19
AAF61014/C
ID AAF61014 standard; DNA: 2397 BP.
XX
AC AAF61014;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF04125.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (OUTA-) OUTAGEN GMBH.
PA (GBFR-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX WPI: 2001-192469/20.
XX
PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other

```

PT related bacteria -
 XX
 PS Claim 1a: Page 62-64; 158pp; German.
 CC This invention describes novel DNA sequences (1) for specific detection
 CC of *Pseudomonas putida* KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1); (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1), (1)', and their fragments, are used as probes
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (1),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.
 XX
 SQ Sequence 2397 BP; 488 A; 792 C; 708 G; 406 T; 3 other;
 XX
 Query Match 56.0%; Score 14; DB 22; Length 2397;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gcgcagcaggtgca 14
 |||||||||
 Db 723 GCGCAGCAGGTGCA 710
 RESULT 20
 AA239011/c
 ID AA239011 standard; cDNA: 3593 BP.
 XX
 AC AA239011:
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Ese2 coding sequence.
 XX
 KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 XX
 PN WO9955728-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;
 XX
 DR WPI: 2000-052802/04.
 DR P-PSDB: AAY57445.

XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 25; Page 46-48; 99pp; English.
 XX
 CC The present sequence encodes mouse Ese2. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (1) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (1) (or its antagonists, mimetics, fragments and inactive
 CC mutants); (1)-specific antibodies (Ab); sequences antisense to the (1)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (1) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamn to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 SQ Sequence 3593 BP; 1171 A; 783 C; 850 G; 789 T; 0 other;
 XX

Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 agtgcatcgccag 21
 |||||||||
 Db 923 AGGTGCATCGCCAG 910
 RESULT 21
 AAQ81778/c
 ID AAQ81778 standard; DNA: 4182 BP.
 XX
 AC AAQ81778:
 XX
 DT 09-AUG-1995 (first entry)
 XX
 DE DNA encoding cytoadhesin protein.
 XX
 KW Cytoadhesin protein; diagnostic probe; unusual codon usage;
 KW immunoassay; antigen; ds.
 XX
 OS Mycoplasma gallisepticum.
 XX
 Key Location/Qualifiers
 FH 713..4081
 FT /*tag= a
 FT /product= cytoadhesin protein
 FT misc_difference 887..889
 FT /*tag= b
 FT /codon= seq:tga, aa:trp
 FT misc_difference 998..1000
 FT /*tag= c
 FT /codon= seq:tga, aa:trp
 FT misc_difference 1484..1486
 FT /*tag= d
 FT /codon= seq:cta, aa:lys
 FT misc_difference 1487..1489
 FT /*tag= e
 FT /codon= seq:tta, aa:lys
 FT misc_difference 1559..1561

FT		/tag= f	/codon= seq: tta, aa: lys
FT		misc_difference	1592..1594
FT		/tag= g	/codon= seq: tta, aa: lys
FT		misc_difference	1676..1678
FT		/tag= h	/codon= seq: tga, aa: trp
FT		misc_difference	1715..1717
FT		/tag= i	/codon= seq: ctt, aa: lys
FT		misc_difference	1853..1855
FT		/tag= j	/codon= seq: tga, aa: trp
FT		misc_difference	1890..1892
FT		/tag= k	/codon= seq: tga, aa: trp
FT		misc_difference	1904..1906
FT		/tag= l	/codon= seq: tta, aa: lys
FT		misc_difference	1961..1963
FT		/tag= m	/codon= seq: tga, aa: trp
FT		misc_difference	2132..2134
FT		/tag= n	/codon= seq: ctt, aa: lys
FT		misc_difference	2237..2239
FT		/tag= o	/codon= seq: tga, aa: trp
FT		misc_difference	2273..2275
FT		/tag= p	/codon= seq: tga, aa: trp
FT		misc_difference	2417..2419
FT		/tag= q	/codon= seq: tga, aa: trp
FT		misc_difference	2504..2506
FT		/tag= r	/codon= seq: tta, aa: lys
FT		misc_difference	2609..2611
FT		/tag= s	/codon= seq: tga, aa: trp
FT		misc_difference	2639..2641
FT		/tag= t	/codon= seq: tga, aa: trp
FT		misc_difference	2858..2860
FT		/tag= u	/codon= seq: tga, aa: trp
FT		misc_difference	3020..3022
FT		/tag= v	/codon= seq: tga, aa: trp
FT		misc_difference	3083..3085
FT		/tag= w	/codon= seq: tga, aa: trp
FT		misc_difference	3138..3140
FT		/tag= x	/codon= seq: tga, aa: trp
PN		US5378820-A.	
XX		03-JAN-1995.	
PD		09-NOV-1992;	92US-0973257.
XX		09-NOV-1992;	92US-0973257.
PF		09-NOV-1992;	92US-0973257.
XX		09-NOV-1992;	92US-0973257.
PI		DOHMS J E.	
XX		(KEEL/) KEELER C L.	
XX		Döhms JE, Keeler CL.	
XX		WPI: 1995-051314/07.	
XX		P-PsDB; AAR64927.	

PT	Nucleic acid encoding cytohesin protein - used as a probe to
PR	diagnose Mycoplasma gallisepticum infection in poultry
XX	
PS	Claim 1; Column 7-16; 13pp; English.
XX	
CC	The sequence encodes a cytohesin protein from Mycoplasma
CC	gallisepticum. All or part of the sequence may be used as a probe
CC	for diagnosis of M. gallisepticum infection in poultry, e.g. turkey
CC	or fowl. The sequence has several USA codons, which may be
CC	eliminated to avoid problems of expression in vitro due to irregular
CC	codon usage. The sequence may be cloned in a vector for recombinant
CC	cytohesin production, for use as an antigen in immunosay formats.
SQ	
XQ	Sequence 4182 BP; 1456 A; 763 C; 746 G; 1217 T; 0 other;
OY	
Query Match	56.0%; Score 14; DB 16; Length 4182;
Best Local Similarity	100.0%; Pred.No. 77;
Matches 14; Conservative	0; Mismatches 0; Indels 0; Gaps 0
Db	
3 gcaacagatgcac 16	
1260 GCACAGCTGCATC 1247	
RESULT 22	
AAZ39010/C	
ID AAZ39010 standard; cDNA; 4625 BP.	
XX AAZ39010;	
AC	
XX	
DT 28-FEB-2000 (first entry)	
XX	
DE Mouse Eseq2 full length cDNA sequence.	
XX	
KW Mouse; murine; Eseq1; Eseq2; endocytosis; vesicular trafficking;	
KM regulation; actin cytoskeleton; detection; cancer; infection;	
KM EH-domain and SH3-domain regulator of endocytosis; antitumor;	
XX antiproliferative; antiviral; ss.	
XX	
OS Mus sp.	
XX	
PN WO955728-A2.	
XX	
PD 04-NOV-1999.	
XX	
PF 27-APR-1999; 99WO-CA00375.	
XX	
PR 27-APR-1998; 98CA-2230201.	
XX	
PR 05-FEB-1999; 99US-0118739.	
XX	
PA (HSCR-) HSC RES & DEV LP.	
XX	
P1 Egan SE, Wang W, Sengar A:	
XX	
DR WPI: 2000-052802/04.	
XX	
P-PSDB; AA57445.	
XX	
PT New nucleic acid encoding Eseq1 and 2 proteins, involved in regulation	
PT of endocytosis, used e.g. for treating cancer or preventing viral	
XX	
XX Infection -	
XX	
PS Claim 25; Page 43-46; 99pp; English.	
XX	
CC The present sequence encodes mouse Eseq2. The present invention	
CC specifically describes mammalian Eseq1 and 2 proteins (I) and their splice	
CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)	
CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.	
CC Generally (I) (or its (antagonists, mimetics, fragments and inactive	
CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)	
CC polynucleotide; agents that downregulate expression of Ese genes or	
CC antagonists of an Ese binding partner are used to treat diseases	

CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission.

SQ Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4625;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatcgccag 21
 |||||
 Db 1173 AGTGCATCGCCAG 1160

RESULT 23
 AAZ39027/C
 ID AAZ39027 standard; cDNA; 4975 BP.

XX AAZ39027;

XX 28-FEB-2000 (first entry)

XX Mouse Ese2L coding sequence.

XX Mouse; murine; Ese2; endocytosis; vesicular trafficking;

KW regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;

KW antiproliferative; antiviral; ss.

XX Mus sp.

XX WO955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI; 2000-052802/04.

XX P-PSDB; AAY57450.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation

XX of endocytosis, used e.g. for treating cancer or preventing viral

XX infection -

XX Claim 25; Page 66-69; 99pp; English.

XX The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and

CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese2L coding sequence.

SQ Sequence 4975 BP; 1540 A; 1166 C; 1201 G; 1068 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4975;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatcgccag 21
 |||||
 Db 923 AGTGCATCGCCAG 910

RESULT 24
 AAT97296/C
 ID AAT97296 standard; DNA; 5362 BP.

XX AAT97296;

XX 14-MAY-1998 (first entry)

XX Mouse additional sex comb polypeptide encoding DNA.

XX Mouse; additional sex comb; Asx; cancer; anti-proliferative;

KW dysregulation; de-differentiation; oncogene; ss.

XX Mus sp.

XX Key Location/Qualifiers

XX CDS 93..4637

XX FT /tag="a"

XX FT /product="Asx"

XX WO9742213-A1.

XX 13-NOV-1997.

XX 08-MAY-1997; 97WO-US07806.

XX 08-JUL-1996; 96US-0021383.

XX 08-MAY-1996; 96US-0016585.

XX (CHIR) CHIRON CORP.

XX Randazzo F;

XX WPI; 1997-558903/51.

XX P-PSDB; AAW37392.

XX New mammalian additional sex comb polypeptide(s) and related nucleic

XX acid - used for diagnosis of cancer, screening for

XX anti-proliferative compounds and for inducing de-differentiation and

XX dysregulation of cells

XX Example 5; Page 44-48; 64pp; English.

XX The present sequence encodes mouse Asx (additional sex comb)
 CC polypeptide. Overexpression of Asx, or expression of a variant Asx, is
 CC indicative of cancer, especially colorectal adenocarcinoma, carcinoma,
 CC melanoma, lymphoma or leukemia. Diagnosis (or monitoring of disease) is
 CC done using DNA encoding Asx-based probes; amplification with DNA

CC encoding Asx-derived primers; hybridisation with branched DNA (bDNA)
 CC probes that are specific for DNA encoding Asx; using Ab or by
 CC quantitation of RNA. Modulators that decrease transcription of Asx are
 CC potentially useful as antineoplastic agents, e.g. for treating cancers
 CC specified above or other proliferative disorders such as restenosis,
 CC benign prostatic hypertrophy, retinopathy, psoriasis, arthritis and
 CC premalignant lesions. These conditions can also be treated with DNA
 CC encoding Asx in standard gene therapy processes. The Asx gene promoter
 CC is useful for tissue-specific expression (e.g. in treatment of lung or
 CC colorectal cancers) and the untranslated regions can be used as probes.
 CC Also Asx (1) induces de-differentiation by treatment of progenitor cells
 CC and (ii) causes dysregulation of controlled cell growth. Further it is
 CC used to raise Ab (useful as diagnostic immunoassay reagents or
 CC therapeutically for inhibiting DNA encoding Asx activity); as competitor
 CC in immunoassays and for purification of Ab.

SQ Sequence 5362 BP; 1430 A; 1354 C; 1399 G; 1179 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 5362;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcaggtgcac 16
 |||||
 DB 3354 GCACGAGTGCATC 3341

RESULT 25

AAT97294/C
 ID AAT97294 standard; DNA: 5362 BP.

AC AAT97294;

DT 14-MAY-1998 (first entry)

DE Mouse additional sex comb polypeptide encoding DNA.

KM Mouse; additional sex comb; Asx: tumour suppressor; anti-proliferative;
 cancer; differentiation; cell growth; ss.

OS Mus sp.

Key Location/Qualifiers
 FT CDS 93..4637
 FT /*tag= a
 FT /product= "Asx"

PN MO9742219-A1.

PD 13-NOV-1997.

PF 08-MAY-1997; 97MO-US07805.

PR 08-JUL-1996; 96US-0021383.

PR 08-MAY-1996; 96US-0016585.

PA (CHIR) CHIRON CORP.

PI Randazzo F;

DR WPI: 1997-558906/51.

DR P-PSDB; AAM37390.

PT New mammalian additional sex comb proteins and related nucleic acid
 PT - used for diagnosis and treatment of cancer, screening for
 PT anti-proliferative compounds and for inducing cell growth and
 PT differentiation

PS Example 5; Page 39-42; 55pp; English.

CC The present sequence encodes mouse Asx (additional sex comb) protein.
 CC Over- or under-expression of Asx, or expression of a variant Asx, is

CC indicative of cancer, especially colorectal adenocarcinoma, melanoma,
 CC lymphoma or leukaemia. Diagnosis (or monitoring of disease) is done
 CC using DNA encoding Asx-based probes; amplification with DNA encoding
 CC Asx-derived primers; hybridisation with branched DNA (bDNA) probes that
 CC are specific for DNA encoding Asx; using Ab or by quantitation of RNA.
 CC Modulators that increase transcription of Asx are potentially useful as
 CC antineoplastic agents, e.g. for treating cancers specified above or
 CC other proliferative disorders such as restenosis, benign prostatic
 CC hypertrophy, retinopathy, psoriasis, arthritis and premalignant lesions.
 CC These conditions can also be treated with DNA encoding Asx in standard
 CC gene therapy processes. The Asx gene promoter is useful for tissue-
 CC specific expression (e.g. in treatment of lung or colorectal cancers)
 CC and the untranslated regions can be used as probes. Also Asx is used (1)
 CC to treat neoplasia; (ii) to induce differentiation by treatment of
 CC progenitor cells and (iii) to regulate growth in cells where growth is
 CC dysregulated. Further it is used to raise Ab (useful as diagnostic
 CC immunoassay reagents or therapeutically for inhibiting Asx activity); as
 CC competitor in immunoassays and for purification of Ab.

SQ Sequence 5362 BP; 1430 A; 1354 C; 1399 G; 1179 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 5362;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcaggtgcac 16
 |||||
 DB 3354 GCACGAGTGCATC 3341

RESULT 26

AAV21186/C
 ID AAV21186 standard; DNA: 5676 BP.

AC AAV21186;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.

KM Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 polyketide synthase; actinomycete; ansamycin; ds.

OS Amycolatopsis mediterranei.

Key Location/Qualifiers
 FT CDS 3..5676
 FT /*tag= a
 FT /product= "polyketide synthase"
 FT /note= "no stop codon given"

PN MO9807868-A1.

PD 26-FEB-1998.

PF 18-AUG-1997; 97MO-EP04495.

PR 20-AUG-1996; 96EP-0810551.

PA (NOVS) NOVARTIS AG.

PI Engel N, Schupp T, Toupet C;

DR WPI: 1998-169172/15.

DR P-PSDB; AAM52844.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues

PS Claim 4; Page 38-43; 205pp; English.

CC The present sequence represents Amycolatopsis mediterranei strain wt3136

CC 5.7 kb KpnI fragment DNA, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polypeptides.
CC synthases, which can be used for assembling a library of polypeptides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX
SQ Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5676;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcacgcga 20
|||
DB 3923 CAGGTGCATCGCCA 3910

RESULT 27
AAZ39026/C
ID AAZ39026 standard; cDNA; 6014 BP.

XX
AC AAZ39026;
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse Ese2L cDNA sequence.
XX
KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
OS Mus sp.
XX
PN W09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Wang W, Sengar A;
XX
DR WPI: 2000-052802/04.
DR P-PSDB: AAY57450.
XX
PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
PS Claim 25; Page 63-66; 99pp; English.

XX
CC The present invention specifically describes mammalian Ese1 and 2
CC proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-

CC mediated endocytosis (as a complex with Esp15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Ese genes or antagonists of an Ese binding partner are
CC used to treat diseases associated with undesirable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Ese mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
CC dynamin to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse Ese2L cDNA sequence.

XX
SQ Sequence 6014 BP; 1833 A; 1389 C; 1454 G; 1338 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 6014;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcacgcga 21
|||
DB 1173 AGGTGCATCGCCAG 1160

RESULT 28
AAA81733/C
ID AAA81733 standard; DNA; 12893 BP.

XX
AC AAA81733;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gmm_280 SEQ ID NO:280.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN W0200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rapunoli R, Pizzi M;
XX
DR WPI: 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisseria infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 1561-1565; 1760pp; English.

XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA

CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisseria bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SQ Sequence 12893 BP; 2829 A; 3164 C; 3741 G; 3159 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 12893;
 Best local Similarity 100.0%; Pred. No. 76;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagcagcagtgatc 14
 |||||
 Db 3743 CCGCAGCAGGTGCA 3730

RESULT 29
 AAV62159
 ID AAV62159 standard; DNA; 20387 BP.
 XX
 AC AAV62159;
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
 XX
 KM HSV-2 strain SB5; immunological response induction; therapy;
 XX antiviral identification; viral protein inhibitor; SS.
 OS
 XX Herpes simplex virus type 2.
 FH
 FH Key Location/Qualifiers
 FT complement (1..772)
 FT /tag- a
 FT /transl_except= (pos: 1, aa: Xaa-Xaa-Xaa-Xaa-Xaa)
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72114;
 FT Xaa = unspecified amino acid"
 FT 802..2073
 FT /tag- b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72115"
 FT complement (2413..3504)
 FT /tag- c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72116"
 FT complement (3748..7938)
 FT /tag- d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72117"
 FT complement (8185..8853)
 FT /tag- e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72118"

FT CDS 9189..11102
 FT /tag- f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in AAW72119"
 FT complement (11323..14091)
 FT /tag- g
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72120"
 FT complement (14176..15306)
 FT /tag- h
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72121"
 FT 14989..16053
 FT /tag- i
 FT /product= "ORF#9 protein"
 FT /note= "encoded protein shown in AAW72122"
 FT 16334..18091
 FT /tag- j
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72123"
 FT 18312..20240
 FT /tag- k
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72124"

XX WO9820016-A1.
 XX 14-MAY-1998.
 XX
 XX 31-OCT-1997; 97WO-US20016.
 XX
 XX 09-JUN-1997; 97US-0049018.
 XX 04-NOV-1996; 96DS-0030279.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;
 XX Esser KM, Leary JF;
 XX
 XX WPI; 1998-286847/25.
 XX F-PSDB; AAW72114, AAW72115, AAW72116, AAW72117, AAW72118, AAW72119,
 XX AAW72120, AAW72121, AAW72122, AAW72123, AAW72124.
 XX

XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 XX and treatment of infection or inducing immunological response in
 XX mammal

XX Claim 1; Page 410-418; 748pp; English.

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
 XX sequence of the invention. This sequence was isolated from HSV-2 strain
 XX SB5 (deposited as ATCC VR-2546), is designated Contig ID 15, and encodes
 XX 11 HSV-2 proteins. The proteins can be used for the treatment or
 XX prevention of disease, to induce an immunological response in a mammal or
 XX to identify inhibitors, activators or novel antivirals. Antagonists of
 XX the proteins can be used to inhibit a viral polypeptide. The DNA sequence
 XX or a vector containing it can also be used to induce an immunological
 XX response in a mammal.

SQ Sequence 20387 BP; 3318 A; 6905 C; 7034 G; 3130 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 20387;
 Best local Similarity 100.0%; Pred. No. 75;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcagcagtgatc 16
 |||||
 Db 16597 gcagcagcagtgatc 16610

RESULT 30
 AAT30681/C

```

ID  AAT30681 standard; DNA; 20710 BP.
XX
AC  AAT30681;
XX
DT  11-SEP-1996 (first entry)
XX
DE  Kaposi's sarcoma associated herpesvirus clone KS5.
XX
KM  Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;
XX  therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.
OS  Kaposi's sarcoma associated herpesvirus.
XX
FH  location/Qualifiers
FH  complement (1..364)
FT  /tag= a
FT  /note= "incomplete ORF20"
FT  363..2105
FT  /tag= b
FT  /product= thymidine kinase
FT  /note= "ORF21"
FT  2093..4285
FT  /tag= c
FT  /product= glycoprotein H
FT  /note= "ORF22"
FT  complement (4282..5496)
FT  /tag= d
FT  /note= "ORF23"
FT  complement (5500..7758)
FT  /tag= e
FT  /note= "ORF24"
FT  7757..11887
FT  /tag= f
FT  /note= "ORF25"
FT  11913..12830
FT  /tag= g
FT  /product= virion polypeptide VP23
FT  /note= "ORF26"
FT  12851..13723
FT  /tag= h
FT  /product= minor capsid protein
FT  /note= "ORF27"
FT  13969..14331
FT  /tag= i
FT  /note= "ORF28"
FT  complement (14343..15485)
FT  /tag= j
FT  /product= (with ORF29A) putative DNA packaging
FT  protein
FT  /note= "ORF29B"
FT  15609..15842
FT  /tag= k
FT  /note= "ORF30"
FT  15839..16423
FT  /tag= l
FT  /note= "ORF31"
FT  16390..17754
FT  /tag= m
FT  /note= "ORF32"
FT  17747..18685
FT  /tag= n
FT  /note= "ORF33"
FT  complement (18743..19662)
FT  /tag= o
FT  /product= (with ORF29B) putative DNA packaging
FT  protein
FT  /note= "ORF29A"
FT  19661..20644
FT  /tag= p
FT  /note= "ORF34"
FT  20625..20710
FT  /tag= q
FT  /note= "incomplete ORF35"

```

```

XX  XX
XX  W09615779-A1.
XX  30-MAY-1996.
XX  21-NOV-1995; 95WO-US15138.
XX  11-APR-1995; 95US-0420235.
XX  21-NOV-1994; 94US-0343101.
XX  (UYCO ) UNIV COLUMBIA NEW YORK.
XX  Chang Y, Moore PS;
XX  WPI: 1996-268320/27.
XX  Herpes virus associated with Kaposi's sarcoma - also definitive DNA
XX  sequences, useful for diagnosis of and to develop prods. for
XX  treatment of Kaposi's sarcoma
XX  Claim 16; Page 164-175; 277pp; English.
XX  Phage lambda clone KS5 (AAT30681) was isolated from an AIDS-Kaposi's
XX  sarcoma (KS) genomic DNA library by screening with a specific DNA,
XX  KS330Bam, found in KS lesions. Sequence analysis of KS5 showed
XX  that KS330Bam is part of a larger herpesvirus genome. This agent
XX  was defined as a human gamma-2 herpesvirus that can be continuously
XX  cultured in naturally-transformed, EBV-coinfected lymphocytes from
XX  AIDS-related body-cavity based lymphomas. KS5 has 17 protein-coding
XX  regions (see also AAT30682-98) identified by homology to known gamma-
XX  herpesvirus sequences. Probes that bind specifically to the KS
XX  associated herpesvirus can be used for KS diagnosis. Antisense or
XX  triplex oligonucleotides are useful for prophylaxis or treatment of
XX  KS, and the peptide products (see also AA97830-46) of the open reading
XX  frames are useful as vaccines.
XX  Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 17; Length 20710;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcacacat 24
    |||||
Db 2138 TGCATGCCGACCAT 2125

RESULT 31
AAT16806/c
ID AAT16806 standard; DNA; 20710 BP.
XX
AC AAT16806;
XX
DT 13-AUG-1996 (first entry)
XX
DE Kaposi's sarcoma associated herpesvirus clone KS5.
XX
KM Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
XX  ss.
XX
OS Kaposi's sarcoma associated herpesvirus.
XX
PN W09606159-A1.
XX
XX 29-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10194.
XX
XX 11-APR-1995; 95US-0420235.
XX 18-AUG-1994; 94US-0292365.
XX 21-NOV-1994; 94US-0343100.
XX

```

PA (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRAN/) GRANT D E.
 PA (VIE/L/) VIELE L.
 XX
 FI Chang Y, Moore PS, Grant DE, Viele L;
 XX
 DR WPI: 1996-151362/15.
 XX
 PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated
 PT vectors and proteins, used in detection and vaccination.
 XX
 PS Claim 16, Page 177-188, 305pp. English.
 PS
 CC Representational difference analysis was used to identify unique
 CC DNA sequences in Kaposi's sarcoma (KS) tissue. An isolated sequence,
 CC K5330bam (AA16824), was used to screen a KS lesion genomic library. A
 CC lambda phage, K55, was identified and the sequence of its insert
 CC (AA16808) was detd. This represents a novel infectious agent
 CC associated with AIDS-KS, Kaposi's sarcoma associated herpesvirus (KSHV).
 CC The K55 fragment has 15 complete ORFs and 2 incomplete ORFs (see also
 CC AA16807-23). Portions of the sequence may be incorporated into a
 CC vector and expressed in host cells to produce peptides (see also
 CC AAR93601-17) useful in vaccines or for raising antibodies. The DNA may
 CC itself be used to immunise a subject, or used to design therapeutic
 CC antisense and triplex molecules or diagnostic probes, or to raise
 CC transgenic animals.
 CC
 XX Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;
 S0
 Query Match 56.0%; Score 14; DB 17; Length 20710;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 tgcacgcacacat 24
 ||||||||||||
 DB 2138 TGCATCGCCAGCAT 2125
 RESULT 32
 AAV62134
 ID AAV62134 standard; DNA; 26338 BP.
 XX
 AC AAV62134;
 XX
 DT 07-DEC-1998 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 104 DNA sequence.
 XX
 XX HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 XX Location/Qualifiers
 FH complement (256..3021)
 FT
 FT CDS
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72033"
 FT complement (2895..3908)
 FT
 FT CDS
 FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72034"
 FT complement (3637..6093)
 FT
 FT CDS
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72035"
 FT complement (6263..6745)
 FT
 FT CDS
 FT /tag= d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72036"
 FT 6775..8046
 FT /tag= e

FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72037"
 FT complement (8386..9477)
 FT
 FT CDS
 FT /tag= f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in AAW72038"
 FT complement (9721..13911)
 FT
 FT CDS
 FT /tag= g
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72039"
 FT complement (14158..14826)
 FT
 FT CDS
 FT /tag= h
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72040"
 FT 15162..17075
 FT
 FT CDS
 FT /tag= i
 FT /product= "ORF#9 protein"
 FT /note= "encoded protein shown in AAW72041"
 FT complement (17295..19110)
 FT
 FT CDS
 FT /tag= j
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72042"
 FT complement (19111..20064)
 FT
 FT CDS
 FT /tag= k
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72043"
 FT complement (20149..21279)
 FT
 FT CDS
 FT /tag= l
 FT /product= "ORF#12 protein"
 FT /note= "encoded protein shown in AAW72044"
 FT 20962..22026
 FT
 FT CDS
 FT /tag= m
 FT /product= "ORF#13 protein"
 FT /note= "encoded protein shown in AAW72045"
 FT 22307..23887
 FT
 FT CDS
 FT /tag= n
 FT /product= "ORF#14 protein"
 FT /note= "encoded protein shown in AAW72046"
 FT 23778..24065
 FT
 FT CDS
 FT /tag= o
 FT /product= "ORF#15 protein"
 FT /note= "encoded protein shown in AAW72047"
 FT 24286..25632
 FT
 FT CDS
 FT /tag= p
 FT /product= "ORF#16 protein"
 FT /note= "encoded protein shown in AAW72048"
 FT 25457..26215
 FT
 FT CDS
 FT /tag= q
 FT /product= "ORF#17 protein"
 FT /note= "encoded protein shown in AAW72049"
 FT
 XX
 PN MO9820016-A1.
 XX
 PD 14-MAY-1998.
 XX
 PD 31-OCT-1997; 97WO-US20016.
 XX
 PF 09-JUN-1997; 97US-0049018.
 PR 04-NOV-1996; 96US-0030279.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX
 PI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
 PI Esser KM, Leaty JJ;
 XX
 DR WPI: 1998-286847/25.
 DR P-RSDB: AAW72033, AAW72034, AAW72035, AAW72036, AAW72037, AAW72038,
 DR AAW72039, AAW72040, AAW72041, AAW72042, AAW72043, AAW72044, AAW72045,
 DR AAW72046, AAW72047, AAW72048, AAW72049.
 XX
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal

XX
PS Claim 1: Page 226-237; 748pp; English.
XX
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC sequence of the invention. This sequence was isolated from HSV-2 strain
CC SB5 (deposited at ATCC VR-2546), is designated Contig ID 104, and encodes
CC 17 HSV-2 proteins. The proteins can be used for the treatment or
CC prevention of disease, to induce an immunological response in a mammal or
CC to identify inhibitors, activators or novel antivirals. Antagonists of
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC or a vector containing it can also be used to induce an immunological
CC response in a mammal.
XX
SQ Sequence 26338 BP; 4257 A; 8879 C; 9241 G; 3959 T; 2 other;

Query Match 56.0%; Score 14; DB 19; Length 26338;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcac 16
DB 22570 gcagcaggtgcac 22583

RESULT 33
AAV73803/C
ID AAV73803 standard; DNA; 35100 BP.
XX
AC AAV73803;
XX
DT 25-FEB-1999 (first entry)
XX
DE KSHV LTR DNA (nucleotides 35,101-70,200).
XX
XX Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
XX dihydrofolate reductase; LTR; long unique region; vaccine; prophylaxis;
XX diagnosis; treatment; HHV8; thymidine kinase; glycoprotein H;
XX capsid protein; packaging protein; helicase primase; glycoprotein M;
XX viral protein kinase; alkaline exonuclease; virion assembly protein;
XX uracil DNA glycosylase; UDG; glycoprotein L; ds.
XX
OS Kaposi's sarcoma-associated herpesvirus.
XX
XX US5849564-A.
XX
XX 15-DEC-1998.
XX
XX 29-NOV-1996; 96US-0770379.
XX
XX 29-NOV-1996; 96US-0770379.
XX
XX (UYCO) UNITV COLUMBIA NEW YORK.
XX
XX Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX
XX WPI; 1999-069741/06.
XX
XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
XX dihydrofolate reductase and is useful for treatment, prophylaxis
XX or diagnosis of Kaposi's sarcoma
XX
XX Disclosure: Column 97-126; 109pp; English.
XX
XX This sequence is a fragment of the Kaposi's sarcoma-associated
XX herpesvirus (KSHV) LTR (long unique region). This fragment contains
XX coding regions for ORF20, ORF21 which encodes thymidine kinase, ORF22
XX which encodes glycoprotein H, ORF23, ORF24, ORF25 which encodes the
XX major capsid protein, ORF26 which encodes capsid protein II, ORF27,
XX ORF28, ORF29b which encodes packaging protein II, ORF30, ORF31, ORF32,
XX ORF33, ORF29a which encodes packaging protein I, ORF34, ORF35, ORF36
XX which encodes viral protein kinase, ORF37 which encodes alkaline
XX exonuclease, ORF38, ORF39 which encodes glycoprotein M, ORF40 which

CC encodes helicase subunit I, ORF41 which encodes helicase primase
CC subunit II, ORF42, ORF43 which encodes capsid protein III, ORF44 which
CC encodes helicase primase subunit III, ORF45 which encodes the virion
CC assembly protein, ORF46 which encodes uracil DNA glycosylase (UDG),
CC ORF47 which encodes glycoprotein L. KSHV is a new human Herpesvirus
CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
CC form of neoplasm occurring in persons with acquired immune deficiency
CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
CC diagnosis and treatment of a subject with Kaposi's sarcoma and for
CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a
CC cell.
XX
SQ Sequence 35100 BP; 8016 A; 9231 C; 9477 G; 8376 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 35100;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgccagcat 24
DB 2058 TGCAATGCCAGCAT 2045

RESULT 34
AAV21187/C
ID AAV21187 standard; DNA; 53789 BP.
XX
AC AAV21187;
XX
DT 24-JUL-1998 (first entry)
XX
DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
XX
XX OS
XX FH Key
XX FT Location/Qualifiers
XX FT 1825..15543
XX FT /*tag= a
XX FT /label= ORF_A
XX FT /product= "polyketide synthase"
XX FT 15550..30759
XX FT /*tag= b
XX FT /label= ORF_B
XX FT /product= "polyketide synthase"
XX FT 30895..36060
XX FT /*tag= c
XX FT /label= ORF_C
XX FT /product= "polyketide synthase"
XX FT 36259..41325
XX FT /*tag= d
XX FT /label= ORF_D
XX FT /product= "polyketide synthase"
XX FT 41373..51614
XX FT /*tag= e
XX FT /label= ORF_E
XX FT /product= "polyketide synthase"
XX FT 51713..5293
XX FT /*tag= f
XX FT /label= ORF_F
XX FT /product= "polyketide synthase"
XX
XX M09807868-A1.
XX
XX 26-FEB-1998.
XX
XX 18-AUG-1997; 97WO-EP04495.
XX
XX 20-AUG-1996; 96EP-0810551.
XX

PA (NOVS) NOVARTIS AG.
 XX
 PI Engel N, Schupp T, Toupet C;
 XX
 DR WPI: 1998-169172/15,
 DR P-PSDB: AAW52845-W52850.
 XX
 PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues
 XX
 PS Claim 4; Page 53-102; 205pp; English.
 XX
 CC The present sequence represents a Amycolatopsis mediterranei rifamycin
 CC synthesis gene cluster DNA fragment from the present invention. The
 CC DNA fragment comprises a DNA region involved directly or indirectly
 CC in the gene cluster responsible for rifamycin synthesis, including
 CC the adjacent DNA regions to the right and left which, by reason of
 CC their function in connection with rifamycin biosynthesis, qualify
 CC as constituents of this rifamycin gene cluster, and functional
 CC fragments, derivatives or constituents of these. The Amycolatopsis
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
 CC for producing rifamycin, rifamycin analogues or precursors. It can also
 CC be used for inactivating or modifying genes involved in ansamycin or
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant
 CC actinomycetes strains from which the natural rifamycin or ansamycin
 CC biosynthesis gene cluster has been partly or completely deleted. The
 CC DNA fragment can be used for assembling a library of polypeptides.
 CC synthases, which can be used for assembling a library of polypeptides.
 CC A hybridisation probe of the invention can be used for identifying DNA
 CC fragments involved in the biosynthesis of ansamycins.
 XX
 SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
 Query Match 56.0%; Score 14; DB 19; Length 53789;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 caggtgcacgcaca 20
 |||||
 Db 47015 CAGGTGCATCGCCA 47002
 RESULT 35
 AAV62176
 ID AAV62176 standard; DNA; 117213 BP.
 XX
 AC AAV62176;
 XX
 DT 13-JAN-1999 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
 XX
 KW HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 FH Key
 FT Location/Qualifiers
 FT 755..1297
 FT CDS
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72170"
 FT 1170..2174
 FT CDS
 FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72171"
 FT 2229..2930
 FT CDS
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72172"
 FT complement (3130..3735)
 FT /tag= q
 FT CDS

FT
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72173"
 FT complement (3802..6447)
 FT CDS
 FT /tag= e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72174"
 FT 6017..8482
 FT CDS
 FT /tag= f
 FT /product= "ORF#6g protein"
 FT /note= "encoded protein shown in AAW72250"
 FT 6026..8482
 FT CDS
 FT /tag= g
 FT /product= "ORF#6f protein"
 FT /note= "encoded protein shown in AAW72249"
 FT 6065..8482
 FT CDS
 FT /tag= h
 FT /product= "ORF#6e protein"
 FT /note= "encoded protein shown in AAW72248"
 FT 6167..8482
 FT CDS
 FT /tag= i
 FT /product= "ORF#6d protein"
 FT /note= "encoded protein shown in AAW72247"
 FT 6296..8482
 FT CDS
 FT /tag= j
 FT /product= "ORF#6c protein"
 FT /note= "encoded protein shown in AAW72246"
 FT 6326..8482
 FT CDS
 FT /tag= k
 FT /product= "ORF#6b protein"
 FT /note= "encoded protein shown in AAW72245"
 FT 6446..8482
 FT CDS
 FT /tag= l
 FT /product= "ORF#6a protein"
 FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
 FT /transl_except= (pos: 7481..7486, aa: Ile)
 FT /note= "encoded protein shown in AAW72244"
 FT 8457..9347
 FT CDS
 FT /tag= m
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72175"
 FT complement (9604..11855)
 FT CDS
 FT /tag= n
 FT /transl_except= (pos: 11635..11636, aa: Ala)
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72176"
 FT complement (11905..14508)
 FT CDS
 FT /tag= o
 FT /product= "ORF#9b protein"
 FT /note= "encoded protein shown in AAW72222"
 FT complement (11905..14520)
 FT CDS
 FT /tag= p
 FT /product= "ORF#9a protein"
 FT /note= "encoded protein shown in AAW72223"
 FT 14399..15802
 FT CDS
 FT /tag= q
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72177"
 FT complement (15996..16286)
 FT CDS
 FT /tag= r
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72178"
 FT complement (16202..18064)
 FT CDS
 FT /tag= s
 FT /product= "ORF#12 protein"
 FT /note= "encoded protein shown in AAW72179"
 FT complement (18105..19661)
 FT CDS
 FT /tag= t
 FT /product= "ORF#13 protein"
 FT /note= "encoded protein shown in AAW72180"
 FT complement (19415..20074)
 FT CDS
 FT /tag= u
 FT /product= "ORF#14 protein"
 FT /note= "encoded protein shown in AAW72181"
 FT


```

FT CDS 2015..21453
FT /tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in AAW72182"
FT CDS complement (21326..22291)
FT /tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in AAW72183"
FT CDS complement (22546..24654)
FT /tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in AAW72184"
FT CDS 24684..25955
FT /tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in AAW72185"
FT CDS complement (26295..27251)
FT /tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in AAW72186"
FT CDS complement (27630..31754)
FT /tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in AAW72225"
FT CDS complement (27630..31784)
FT /tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in AAW72224"
FT CDS complement (32067..32735)
FT /tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in AAW72187"
FT CDS 33140..34984
FT /tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in AAW72226"
FT CDS 33386..34984
FT /tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in AAW72227"
FT CDS complement (35205..37721)
FT /tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in AAW72188"
FT CDS complement (38058..39188)
FT /tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in AAW72189"
FT CDS 39090..39935
FT /tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT CDS 40216..41973
FT /tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT CDS 42206..44178
FT /tag= aj
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT CDS complement (44853..47297)
FT /tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT CDS 47122..47338
FT /tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in AAW72194"
FT CDS complement (47305..49662)
FT /tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in AAW72195"

```

```

FT CDS complement (50035..51666)
FT /tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in AAW72196"
FT CDS complement (51701..53575)
FT /tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in AAW72197"
FT CDS 54393..58115
FT /tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in AAW72198"
FT CDS complement (58060..58977)
FT /tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in AAW72199"
FT CDS complement (58970..60760)
FT /tag= ar
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in AAW72200"
FT CDS 60759..61151
FT /tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in AAW72201"
FT CDS 61241..62071
FT /tag= at

Query Match 56.0%; Score 14; DB 19; Length 117213;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcac 16
DB 40479 gcagcaggtgcac 40492

RESUL 36
AAV19941/C
ID AAV19941 standard; DNA; 137507 BP.
AC AAV19941;
XX
XX 03-AUG-1998 (first entry)
DE KSHV long unique coding region and terminal repeat.
XX KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
XX interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
XX complement-binding protein; glycoprotein; capsid protein IV; infection;
XX immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
XX lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
XX HIV immune status; anti-inflammatory agent; therapy; ds.
OS Kaposi's sarcoma-associated herpes virus.
XX
FH Key 1142..2794
FT Location/Qualifiers
FT /tag= a
FT /product= complement-binding protein
FT 8699..11236
FT /tag= b
FT /product= glycoprotein B
FT complement (17261..17875)
FT /tag= c
FT /product= interleukin 6
FT complement (21548..21832)
FT /tag= d
FT /product= macrophage inflammatory protein II
FT complement (27137..27424)
FT /tag= e
FT /product= interferon regulatory factor 1
FT 28661..29741
FT /tag= f

```

```

FT      /product= protein T1.1
FT      complement (58976..60175)
FT      /tag= g
FT      /product= glycoprotein M
FT      complement (69412..69915)
FT      /tag= h
FT      /product= glycoprotein L
FT      complement (88410..88910)
FT      /tag= i
FT      /product= interferon regulatory factor 2
FT      89600..90541
FT      /tag= j
FT      /product= interferon regulatory factor 3
FT      90173..90643
FT      /tag= k
FT      /product= glycoprotein X
FT      complement (93636..94127)
FT      /tag= l
FT      /product= interferon regulatory factor 4
FT      complement (111931..112443)
FT      /tag= m
FT      /product= capsid protein IV
FT      complement (123808..127296)
FT      /tag= n
FT      /product= immediate early protein

```

```

PN      WO9804576-A1.
XX
XX      05-FEB-1998.
PD
XX
XX      22-JUL-1997; 97WO-US13346.
PF
XX      29-NOV-1996; 96US-0757669.
PR      25-JUL-1996; 96US-0686243.
PR      25-JUL-1996; 96US-0686349.
PR      25-JUL-1996; 96US-0686350.
PR      25-JUL-1996; 96US-0687253.
PR      25-JUL-1996; 96US-0688814.
PR      05-SEP-1996; 96US-0708678.
PR      10-OCT-1996; 96US-0728323.
PR      13-NOV-1996; 96US-0747887.
PR      13-NOV-1996; 96US-0748640.
XX
XX      (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX      Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX      WPI: 1998-130615/12.
XX
XX      New nucleic acid encoding Kaposi's sarcoma associated herpes virus
XX      proteins - useful for, e.g. detecting levels of HHV8 in, and
XX      preparation of vaccines for treatment of, HIV patients
XX
XX      Example 2: Page 135-203; 230pp; English.
XX
XX      This sequence represents the long unique region and terminal repeat of
XX      the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
XX      as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
XX      invention which encode KSHV polypeptides selected from: (a) viral
XX      macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
XX      (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
XX      (e) capsid protein IV encoded by ORF65; and (e) immediate early protein
XX      encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
XX      by 1I, and antibodies (Ab) specific for the proteins are useful for
XX      detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
XX      fluids or tissue samples. HHV8 infections can be treated with antisease
XX      or triplex forming molecules or agents that bind specifically to the
XX      protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
XX      while the protein can be used in protective vaccines. Ab may also be used
XX      to differentiate between lymphomas, and HHV8 may be implicated in many
XX      other lymphoproliferative diseases such as lymphomas, leukaemia,
XX      CC splenomegaly and mycosis fungoides. Cells and animals containing the
XX      nucleic acid are useful for drug screening. HHV8-derived peptides can be

```

```

CC      CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC      CC can be inhibited with methotrexate. These can also be used to determine
CC      CC the immune status of a patient infected with HIV. HHV8 derived protein
CC      CC viral MIP II may be used as an anti-inflammatory agent for,
CC      CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC      CC 81 open reading frames.
XX
XX      SQ      Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;
XX
XX      Query Match      56.0%; Score 14; DB 19; Length 137507;
XX      Best Local Similarity 100.0%; Pred. No. 72;
XX      Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
XX      QY      11 tgcattgcacagat 24
XX      Db      37158 TGCATTCGACAGCAT 37145
XX
XX      RESULT 37
XX      ID      AAF21608/c
XX      AAF21608 standard; DNA; 349980 BP.
XX
XX      AC      AAF21608;
XX
XX      DT      13-MAR-2001 (first entry)
XX
XX      DE      Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.
XX
XX      KW      Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX      KW      diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX      ds.
XX
XX      OS      Neisseria meningitidis.
XX
XX      PN      WO200006791-A1.
XX
XX      PD      09-NOV-2000.
XX
XX      PE      08-MAR-2000; 2000WO-US05928.
XX
XX      PR      30-APR-1999; 99US-0132068.
XX      08-OCT-1999; 99WO-US23573.
XX      28-FEB-2000; 2000CB-0004695.
XX
XX      PA      (CHIR ) CHIRON CORP.
XX      (GENO-) INST GENOMIC RES.
XX
XX      PI      Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
XX      PI      Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
XX      PI      Frazer CM, Grandi G.
XX      WPI: 2000-647603/62.
XX
XX      PT      Neisseria meningitidis B full length genome sequence and open reading
XX      PT      frames are used to detect, treat and prevent Neisserial infections -
XX      XX
XX      Claim 7: Appendix A: 692pp; English.
XX
XX      The present invention describes the full length genome of
XX      CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
XX      CC to AAF21613 represent fragments of the NMB genomic sequence, as the
XX      CC sequence was too long to go in a record on its own it was split into 8
XX      CC sequences which overlap each other at the beginning and end of each
XX      CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
XX      CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
XX      CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
XX      CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
XX      CC AAF21606 represent PCR primers which are used in the exemplification of
XX      CC the present invention. The NMB genome and fragments from it have
XX      CC antibacterial activity, and can be used in vaccines and gene therapy.
XX      CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
XX      CC proteins can be used in compositions for treating or preventing infection

```

CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.

XX Sequence 349980 BP: 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgca 14
 |||
 Db 141640 GCGCAGCAGTCA 141627

RESULT 38

AAQ41769
 ID AAQ41769 standard; DNA: 21 BP.

XX AAQ41769;

XX 03-SEP-1993 (first entry)

XX Myc EMSA probe production primer MLC-A.

XX Myc: c-myc; mammalian; E box; cancer; therapy; C1; C2; C2'; complex;

XX homo-oligomer; hetero-oligomer; myogenin; Max; oncoprotein; primer;

XX probe; electrophoretic mobility shift assay; EMSA; ss.

XX Synthetic.

FT Key Location/Qualifiers
 FT protein_bind 13..18
 FT /tag- a
 FT /note- "E-Box core sequence"

XX WO9308701-A.

XX 13-MAY-1993.

XX 09-OCT-1992; 92WO-US08603.

XX 30-OCT-1991; 91US-0785567.

XX (GEHO) GEN HOSPITAL CORP.

XX Kingston RE, Papoulas O;

XX WPI; 1993-167291/20.

XX Prod. of c-myc protein from mammalian cells - and detection of c-

XX Myc inhibitors for use in cancer therapy

XX Disclosure; Page 37; 101pp; English.

XX The sequences given in AAQ41764-75 are primers which were used to create

XX double stranded probes for use in an electrophoretic mobility shift

XX assay (EMSA) to isolate Myc binding sites from random sequences. The

XX isolated sequences contain the central E box core of CACGTG which

XX binds very weakly with Myc homo-oligomers (C1 complex), but more

XX tightly with Myc hetero-oligomers (C2 complex). The C2 complex

XX requires a 26-29 kd factor in addition to Myc. The additional factor

XX copurifies with Myc and resembles Max protein. A second copurifying

XX 40-50 kd factor has been identified (forming C2' complex). Sites

XX selected by the C2' complex contain the core CACGTG which bears

XX remarkable homology to a myogenin binding site (see AAQ41763).

XX Oligonucleotides containing the E box can be used in the purification

CC of Myc from a mammalian source. See also AAQ41761-861. The isolated
 CC target sequences may be used in a method to inhibit c-myc oncoprotein
 CC activity.

XX Sequence 21 BP: 4 A; 6 C; 8 G; 3 T; 0 other;

Query Match 52.0%; Score 13; DB 14; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2,8e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gcgcagcaggtgca 14
 |||
 Db 8 gcgcagcaggtgca 20

RESULT 39

AAAX3237
 ID AAAX3237 standard; DNA: 21 BP.

XX AAAX3237;

XX 25-JUN-1999 (first entry)

XX Wheat viviparous 1 (tavp1) primer #4.

XX Wheat; oat; viviparous 1; VPI; afVPI; tavp1; maize; detection; PHS;

XX pre-harvest sprouting; dormant; germination; crop plant; primer; ss.

XX Synthetic.

XX Triticum aestivum.

XX WO9915667-A1.

XX 01-APR-1999.

XX 18-SEP-1998; 98WO-GB02835.

XX 19-SEP-1997; 97GB-0020060.

XX (PLAN-) PLANT BIOSCIENCE LTD.

XX Flinham JE, Gale MD, Holdsworth MD;

XX WPI; 1999-244424/20.

XX New isolated oat and wheat VPI genes, used, e.g. to impose

XX sufficient dormancy to avoid pre-harvest sprouting

XX Claim 56; Page 89; 120pp; English.

XX The present sequence represents a primer for the wheat viviparous 1 (VPI)

XX gene, which keeps mature seeds dormant and inhibits germination. The

XX present invention describes genes which are homologues of the maize

XX Viviparous 1 gene, obtained from oat Avena fatua and wheat which encode

XX polypeptides designated afVPI and tavp1 respectively. The VPI activity

XX keeps mature seeds dormant and inhibits germination and can be used to

XX maintain or impose sufficient intensity and duration of dormancy to

XX avoid pre-harvest sprouting (PHS) before harvest. The products can be

XX used in the production of transformed crop plants having desirable

XX primary or secondary dormancy, or after-ripening properties, and in

XX particular may be resistant to PHS.

XX Sequence 21 BP: 4 A; 5 C; 9 G; 3 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2,8e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcgcagcaggtgcat 15
 |||
 Db 4 gcgcagcaggtgcat 16

```

RESULT 40
AAV1150/c
ID AAV1150 standard; DNA; 22 BP.
XX
XX AAV1150;
AC
XX 14-JUL-1998 (first entry)
XX
XX Oligonucleotide #3 for nucleic acid assay.
DE
XX Nucleic acid assay; hybridising; probe; target; duplex; detection; ss.
XX
XX Synthetic.
OS
XX DE19633436-A1.
XX
XX 26-FEB-1998.
PD
XX
XX 20-AUG-1996; 96DE-1033436.
PF
XX
XX 20-AUG-1996; 96DE-1033436.
PR
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
PA
XX
XX Bergmann F, Herrmann R, Kobold U;
PI
XX
XX WPI; 1998-146248/14.
DR
XX
XX Nucleic acid assay - based on mass of probe-target duplex
PT
XX
XX Disclosure; Fig 1; 9pp; German.
PS
XX
XX AAV1148-V1151 are oligonucleotides used in a novel nucleic acid assay.
CC
XX This assay involves hybridising a probe to a target sequence of the
CC nucleic acid to form a duplex and degrading the nucleic acid to form
CC a fragment of a defined length containing the duplex, which is detected
CC on the basis of its mass. This method enables different nucleic acids to
CC be detected simultaneously does not require the use of labelled probes.
XX
XX Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;
SQ
Query Match 52.0%; Score 13; DB 19; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 catcgccagcatt 25
Db 18 CATGCCGACGACTT 6
RESULT 41
AAA45502/c
ID AAA45502 standard; cDNA; 286 BP.
XX
XX AAA45502;
AC
XX
XX 21-AUG-2000 (first entry)
DT
XX
XX Human secreted expressed sequence tag SEQ ID NO:2077.
DE
XX
XX Human: mouse; chicken; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemokine; proliferative;
XX immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
XX antitumor; osteoprotective; neuroprotective; nootropic; antiparasitic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

```

```

KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
OS
XX WO200021991-A1.
XX
XX 20-APR-2000.
PD
XX
XX 15-OCT-1999; 99WO-US24206.
PF
XX
XX 15-OCT-1998; 98US-010436.
PR
XX
XX (GEMV ) GENETICS INST INC.
PA
XX
XX Jacobs K, McCoy JM, Lavallee ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Bowman MR;
XX
XX WPI; 2000-317938/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX Claim 1; Page 723; 803pp; English.
XX
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
XX sequence tags (SESTs), isolated from human, mouse, chicken and rat
XX tissue sources. The SESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiparasitic; vulnery; antitumor; osteoprotective; neuroprotective;
XX nootropic; antiparkinsonian; antiparasitic; cerebroprotective;
XX anticonvulsant; and antidepressant. The SESTs can be used for gene
XX therapy and in vaccines. The SESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA45926 to AAA45931 represent linker variants which are given
XX in the exemplification of the present invention.
XX
XX Sequence 286 BP; 44 A; 84 C; 69 G; 89 T; 0 other;
SQ
Query Match 52.0%; Score 13; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 gcaggtgcacgc 18
Db 93 GCAGGTGCATGCC 81
RESULT 42
AA214744/c
ID AA214744 standard; cDNA; 300 BP.
XX
XX AA214744;
AC
XX
XX 12-OCT-1999 (first entry)
DT
XX
XX Human gene expression product cDNA sequence SEQ ID NO:2213.
DE

```

```

XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensics; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9338972-A2.
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1159; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AA212532 to AA217779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other:
XX
XX Query Match 52.0%; Score 13; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 2.7e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 gcatgcacgcat 24
XX |||||||||||||
XX 174 GCATGCCACGCAAT 162
XX
XX RESULT 43
XX AA214726
XX ID AA214726 standard; CDNA: 300 BP.
XX

```

```

AC AA214726;
XX
XX 12-OCT-1999 (first entry)
XX
XX Human gene expression product cDNA sequence SEQ ID NO:2195.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensics; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9338972-A2.
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1155; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AA212532 to AA217779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 80 A; 64 C; 63 G; 93 T; 0 other:
XX
XX Query Match 52.0%; Score 13; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 2.7e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 6 gcagtgcatgcgc 18
XX |||||||||||||
XX 180 gcagtgcatgcgc 192
XX

```

RESULT 44
AAZ12980/c
ID AAZ12980 standard; cDNA: 300 BP.
AC AAZ12980;
XX 12-OCT-1999 (first entry)
XX Human gene expression product cDNA sequence SEQ ID NO:449.
DE
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
OS
XX WO938972-A2.
XX
XX 05-AUG-1999.
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1: Page 764; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 69 A; 89 C; 77 G; 65 T; 0 other;

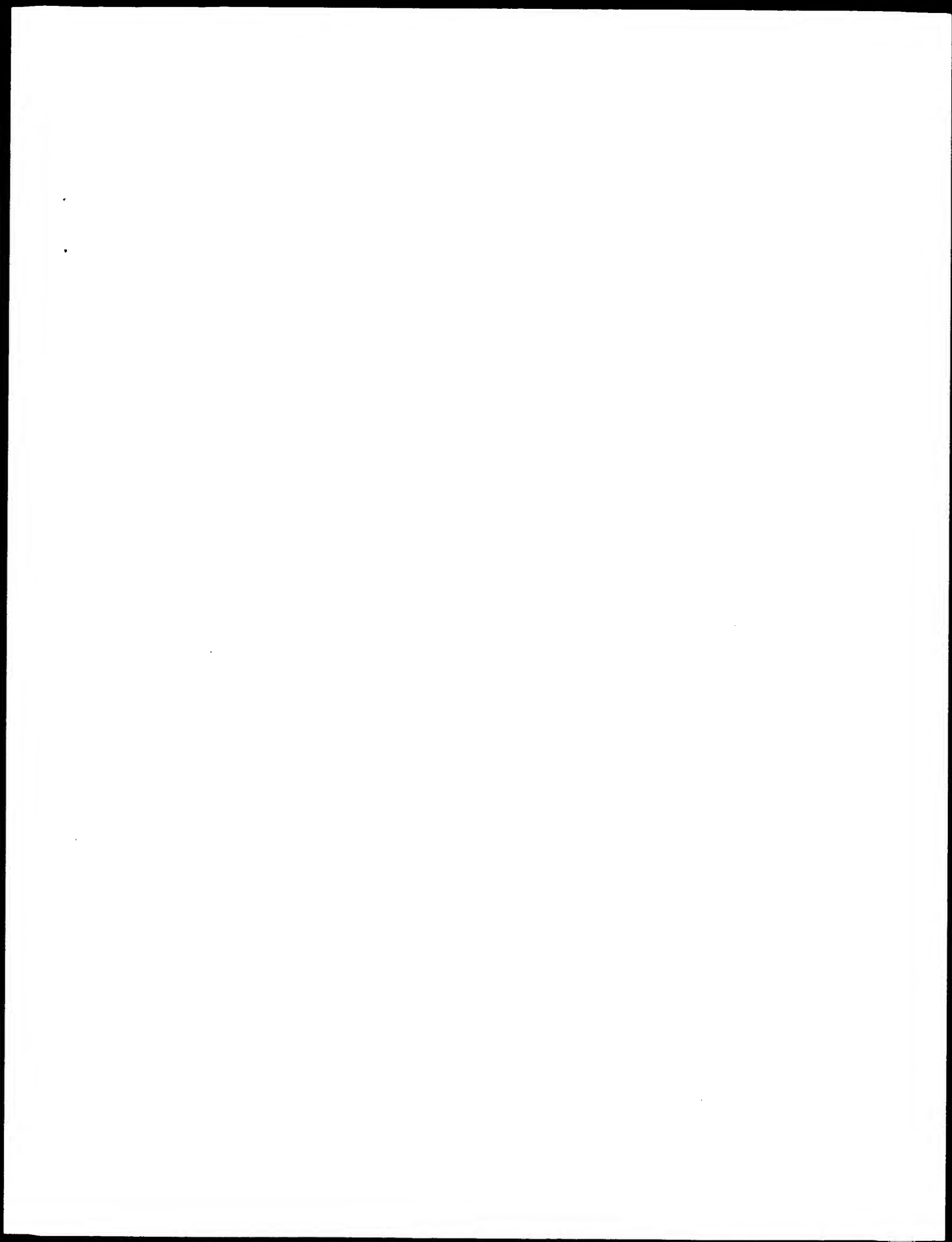
Query Match 52.0%; Score 13; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative, 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ccgacgcaggtgca 14
Db 32 CGCAGCAGGTGCA 20
IIIIIIIIIIII
IIIIIIIIIIII

RESULT 45
AAZ12981/c
ID AAZ12981 standard; cDNA: 300 BP.
AC AAZ12981;
XX 12-OCT-1999 (first entry)
XX Human gene expression product cDNA sequence SEQ ID NO:450.
DE
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
OS
XX WO938972-A2.
XX
XX 05-AUG-1999.
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1: Page 764; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 69 A; 88 C; 77 G; 66 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgcagcaggtgca 14
|||||
Db 32 CGCAGCAGGTGCA 20

Search completed: October 9, 2001, 15:52:25
Job time: 15106 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:27 ; Search time 218.82 seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 ggcgcagcagtgatgcgcgcagcatt 25

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 11705

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appl1
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appl1
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appl1
5	14	56.0	718	4	US-09-142-078-45	Sequence 45, Appl1
6	14	56.0	2193	1	US-08-420-235B-16	Sequence 16, Appl1
7	14	56.0	2193	3	US-08-793-624-16	Sequence 16, Appl1
8	14	56.0	2193	5	PCRT-US95-10194-16	Sequence 16, Appl1
9	14	56.0	2973	2	US-08-343-101A-5	Sequence 5, Appl1
10	14	56.0	2973	3	US-09-183-688-5	Sequence 5, Appl1
11	14	56.0	4182	1	US-07-973-257-1	Sequence 1, Appl1
12	14	56.0	5362	2	US-08-853-310-3	Sequence 3, Appl1
13	14	56.0	20710	1	US-08-420-235B-1	Sequence 1, Appl1
14	14	56.0	20710	3	US-08-793-624-1	Sequence 1, Appl1
15	14	56.0	20710	5	PCRT-US95-10194-1	Sequence 1, Appl1
16	14	56.0	35100	2	US-08-770-379-18	Sequence 18, Appl1
17	14	56.0	35100	4	US-08-757-669A-18	Sequence 18, Appl1
18	14	56.0	304	4	US-09-060-756-726	Sequence 726, App
19	13	52.0	343	4	US-09-060-756-570	Sequence 570, App
20	13	52.0	411	2	US-08-640-978C-2	Sequence 2, Appl1
21	13	52.0	414	2	US-08-640-978C-4	Sequence 4, Appl1
22	13	52.0	447	4	US-08-836-075A-51	Sequence 51, Appl1
23	13	52.0	476	4	US-08-998-416-978	Sequence 978, App
24	13	52.0	1128	1	US-07-872-678A-11	Sequence 11, Appl1
25	13	52.0	1202	2	US-08-428-414A-1	Sequence 1, Appl1
26	13	52.0	1365	4	US-09-319-892-1	Sequence 1, Appl1
27	13	52.0	1624	2	US-07-862-588B-6	Sequence 6, Appl1

28	13	52.0	1775	2	US-07-862-588B-5	Sequence 5, Appl1
29	13	52.0	1875	3	US-09-258-373-21	Sequence 21, Appl1
30	13	52.0	1981	3	US-09-017-706-3	Sequence 3, Appl1
31	13	52.0	1981	3	US-09-017-706-4	Sequence 4, Appl1
32	13	52.0	1981	3	US-09-017-706-5	Sequence 5, Appl1
33	13	52.0	1981	3	US-09-017-706-6	Sequence 6, Appl1
34	13	52.0	1981	3	US-09-017-706-7	Sequence 7, Appl1
35	13	52.0	1981	3	US-09-017-706-8	Sequence 8, Appl1
36	13	52.0	2223	1	US-08-317-522A-8	Sequence 8, Appl1
37	13	52.0	2223	2	US-08-439-818A-8	Sequence 8, Appl1
38	13	52.0	2223	2	US-08-751-965-8	Sequence 8, Appl1
39	13	52.0	2223	2	US-08-738-975-8	Sequence 8, Appl1
40	13	52.0	2223	2	US-08-728-626-8	Sequence 8, Appl1
41	13	52.0	2223	3	US-08-808-599A-8	Sequence 8, Appl1
42	13	52.0	2290	1	US-08-584-226-1	Sequence 1, Appl1
43	13	52.0	2763	3	US-08-941-445A-16	Sequence 16, Appl1
44	13	52.0	3618	1	US-07-872-678A-36	Sequence 36, Appl1
45	13	52.0	3670	1	US-08-386-727-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 5.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgccacatt 25
|||||
DB 79 GCGCAGCAGGTGATCGCCAGCAT 103

RESULT 2

US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SRO ID NO: 7;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 5.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgccacatt 25
|||||
DB 79 GCGCAGCAGGTGATCGCCAGCAT 103

RESULT 3

US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A-15/9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 117..1157
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 117
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "biob"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2295..3050
; OTHER INFORMATION: /codon_start= 2295
; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "bioc"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3750..5039
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3750
; OTHER INFORMATION: /EC_number= 2.6.1.62
; OTHER INFORMATION: /product= "DAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioA"
; OTHER INFORMATION: /number= 5
; OTHER INFORMATION: /standard_name= "S-adenosyl-L-methionine:8-amino-7-oxononanoate
; OTHER INFORMATION: /aminoacids= 1
; NAME/KEY: CDS

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "bioB RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gccgacgagtgatccgacacatt 25
|||||
Db 195 gccgacgagtgatccgacacatt 219
RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SPRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PHO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "bethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "bioD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gccacagctgcatgccagcatt 25
|||||
Db 195 gccacagctgcatgccagcatt 219

RESULT 5
US-09-142-078-45/c
Sequence 45, Application US/09142078
Patent No. 6172041

GENERAL INFORMATION:

APPLICANT: McCabe, R. Tyler

APPLICANT: Zhou, Li-Ming

APPLICANT: Layer, Richard T.

TITLE OF INVENTION: Use of Conantokins

NUMBER OF SEQUENCES: 71

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.

STREET: 555 Thirteenth Street, N.W., Suite 701-E

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/142.078

FILING DATE: 10-FEB-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO US97/12652

FILING DATE: 21-JUL-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/672.377

FILING DATE: 06-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/684.750

FILING DATE: 22-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 2314-135.A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-783-6040

TELEFAX: 202-783-6031

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 718 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (cDNA)

ORIGINAL SOURCE:

ORGANISM: Corvus geographus

FEATURE:

NAME/KEY: CDS

LOCATION: 110..409

US-09-142-078-45

Query Match 56.0%; Score 14; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 caggtgcatgccca 20
|||||
Db 118 CAGGTGATGCCCA 105

RESULT 6
US-08-420-235B-16/c
Sequence 16, Application US/08420235B
Patent No. 5801042

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/420.235B

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 2193 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2193

OTHER INFORMATION:

US-08-420-235B-16

Query Match 56.0%; Score 14; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcagcatt 24
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 7

US-08-793-624-16/c

Sequence 16, Application US/08793624C

Patent No. 6150093

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And

TITLE OF INVENTION: Uses Thereof

FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 56.0%; Score 14; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
DB 46 TGCATGCCACGAT 33

RESULT 8
PCT-US95-10194-16/c

Sequence 16, Application PC/TUS9510194

GENERAL INFORMATION:

APPLICANT: The Trustees of Columbia University in the City of New York

APPLICANT: City

TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS

TITLE OF INVENTION: SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10194

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 2193 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2193

OTHER INFORMATION:

PCT-US95-10194-16

Query Match 56.0%; Score 14; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
DB 46 TGCATGCCACGAT 33

RESULT 9

US-08-343-101A-5/c

Sequence 5, Application US/08343101A

Patent No. 5830759

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma

TITLE OF INVENTION: Virus Sequences And Uses Thereof

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/343,101A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: White Esq., John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0526

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2973 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

US-08-343-101A-5

Query Match 56.0%; Score 14; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
DB 2182 TGCATGCCACGAT 2169

RESULT 10

US-09-183-688-5/c

Sequence 5, Application US/09183688

Patent No. 6093550

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma

TITLE OF INVENTION: Virus Sequences And Uses Thereof

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

```
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-09-183-688-5

Query Match
Best Local Similarity 56.0%; Score 14; DB 3; Length 2973;
Matches 14; Conservative 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccagcat 24
|||||
Db 2182 tgcacgccagcat 2169

RESULT 11
US-07-973-257-1/c
Sequence 1, Application US/07973257
Patent No. 5378620
GENERAL INFORMATION:
APPLICANT: Keeler, Jr. Calvin L.
APPLICANT: Dohms, John E.
TITLE OF INVENTION: Gene Encoding Cytohesin
TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESSES:
ADDRESS: Connolly and Hutz
STREET: 1220 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM/PC or Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/973,257
FILING DATE: 19921109
CLASSIFICATION: 435
PRIOR APPLICATION DATA: No. 5378820e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4182 base pairs
```

```
TYPE: NUCLEIC ACID
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycoplasma Gallisepticum
STRAIN: S6
US-07-973-257-1

Query Match
Best Local Similarity 56.0%; Score 14; DB 1; Length 4182;
Matches 14; Conservative 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgatc 16
|||||
Db 1260 GCACGAGTGATC 1247

RESULT 12
US-08-853-310-3/c
Sequence 3, Application US/08853310
Patent No. 5948640
GENERAL INFORMATION:
APPLICANT: Randazzo, Filippo
TITLE OF INVENTION: Mammalian Additional sex Combs (Asx) Acts as a Tumor Suppre
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/853,310
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Gulth, Joseph H.
REGISTRATION NUMBER: 31,261
REFERENCE/DOCKET NUMBER: 1228.003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3888
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5362 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-853-310-3

Query Match
Best Local Similarity 56.0%; Score 14; DB 2; Length 5362;
Matches 14; Conservative 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgatc 16
|||||
Db 3354 GCACGAGTGATC 3341

RESULT 13
US-08-420-235B-1/c
```

```
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; US-08-420-235B-1

Query Match          56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgacat 24
Db 2138 TGCATGCCACGACAT 2125

RESULT 14
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPM
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO: 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-08-793-624-1
```

```
Query Match          56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgacat 24
Db 2138 TGCATGCCACGACAT 2125

RESULT 15
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10194
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MSC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; PCT-US95-10194-1

Query Match          56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgacat 24
Db 2138 TGCATGCCACGACAT 2125

RESULT 16
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
```

;; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;; NUMBER OF SEQUENCES: 20
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/770,379
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 52342
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-770-379-18

Query Match 56.0%; Score 14; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcatcgccagcat 24
|||||
Db 2058 TGCATCGCCAGCAT 2045

RESULT 17
US-08-757-669A-18/c
;; Sequence 18, Application US/08757669A
;; Patent No. 6183751
;; GENERAL INFORMATION:
;; APPLICANT: Chang, Yuan
;; APPLICANT: Bohenzky, Roy A.
;; APPLICANT: Russo, James J.
;; APPLICANT: Edelman, Isidore S.
;; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 20
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/757,669A
;; FILING DATE:
;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 45185-F
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match 56.0%; Score 14; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcatcgccagcat 24
|||||
Db 2058 TGCATCGCCAGCAT 2045

RESULT 18
US-09-060-756-726
;; Sequence 726, Application US/09060756
;; Patent No. 6183957
;; GENERAL INFORMATION:
;; APPLICANT: Cole, Stewart
;; APPLICANT: Buchrieser-Brosch, Roland
;; APPLICANT: Gordon, Stephen
;; APPLICANT: Billault, Alain
;; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;; FILE REFERENCE: 3495-0169
;; CURRENT APPLICATION NUMBER: US/09/060,756
;; CURRENT FILING DATE: 1998-04-16
;; NUMBER OF SEQ ID NOS: 743
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 726
;; LENGTH: 304
;; TYPE: DNA
;; ORGANISM: Mycobacterium bovis
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (various positions within the sequence)
;; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-726

Query Match 52.0%; Score 13; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcatcgccagc 22
|||||
Db 136 gtgcatcgccagc 148

RESULT 19
US-09-060-756-570
;; Sequence 570, Application US/09060756
;; Patent No. 6183957
;; GENERAL INFORMATION:
;; APPLICANT: Cole, Stewart
;; APPLICANT: Buchrieser-Brosch, Roland
;; APPLICANT: Gordon, Stephen
;; APPLICANT: Billault, Alain
;; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM

;; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;; FILE REFERENCE: 3495-0169
;; CURRENT APPLICATION NUMBER: US/09/060,756
;; CURRENT FILING DATE: 1998-04-16
;; NUMBER OF SEQ ID NOS: 743
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 570
;; LENGTH: 343
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match 52.0%; Score 13; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcacgcacgc 22
|||||
DB 135 gtgcacgcacgc 147

RESULT 20
US-08-640-978C-2/C
; Sequence 2, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISAO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,978C
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C J
; REGISTRATION NUMBER: 32,350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-640-978C-2

Query Match 52.0%; Score 13; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgc 13
|||||
DB 294 GCGCAGCAGGTGC 282

RESULT 21
US-08-640-978C-4/C

; Sequence 4, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISAO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,978C
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C J
; REGISTRATION NUMBER: 32,350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..411
US-08-640-978C-4

Query Match 52.0%; Score 13; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgc 13
|||||
DB 294 GCGCAGCAGGTGC 282

RESULT 22
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS

COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 catgcgcagcatt 25
|||||
Db 406 CATGCAGCAGCATT 418

RESULT 23
US-08-998-416-978/C
Sequence 978, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Rebschunig, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 978:
SEQUENCE CHARACTERISTICS:
LENGTH: 476 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1602UP
US-08-998-416-978

Query Match 52.0%; Score 13; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcaggttc 13
|||||
Db 328 GCGCAGCAGGTTC 316

RESULT 24
US-07-872-678A-11
Sequence 11, Application US/07872678A
Patent No. 5541060
GENERAL INFORMATION:
APPLICANT: Bell, Graeme, et al.
TITLE OF INVENTION: DETECTION OF EARLY-ONSET
TITLE OF INVENTION: NON-INSULIN-DEPENDENT DIABETES MELLITUS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: Post Office Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-07-872-678A-11

Query Match 52.0%; Score 13; DB 1; Length 1128;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 13; Conservative 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgca 14

Db 904 CGCAGCAGGTGCA 916

RESULT 25

US-08-428-414A-1

Sequence 1, Application US/08428414A

Patent No. 5912166

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TITLE OF INVENTION: LEISHMANIASIS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/428,414A

FILING DATE: 21-APR-1995

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Kadlecek, Ann T.

REGISTRATION NUMBER: 39,244

REFERENCE/DOCKET NUMBER: 210121.407

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1202 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 30..998

US-08-428-414A-1

Query Match 52.0%; Score 13; DB 2; Length 1202;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgca 14

Db 133 CGCAGCAGGTGCA 145

RESULT 26

US-09-319-892-1

Sequence 1, Application US/09319892

Patent No. 6177616

GENERAL INFORMATION:

APPLICANT: BARTSCH, Klaus

APPLICANT: KRIETE, Guido

APPLICANT: BROER, Inge

APPLICANT: PUHLER, Alfred

TITLE OF INVENTION: NOVEL GENES CODING FOR AMINO ACID DEACETYLASES WITH

TITLE OF INVENTION: SPECIFICITY FOR N-ACETYL-L-PHOSPHOTRITHICIN, THEIR

TITLE OF INVENTION: ISOLATION AND THEIR USE

FILE REFERENCE: 514412-2005

CURRENT APPLICATION NUMBER: US/09/319,892

CURRENT FILING DATE: 1999-06-14

EARLIER APPLICATION NUMBER: PCT/EP97/06755

EARLIER FILING DATE: 1997-12-03

EARLIER APPLICATION NUMBER: 19652284.6

EARLIER FILING DATE: 1996-12-16

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 1365

TYPE: DNA

ORGANISM: Stenotrophomonas maltophilia

US-09-319-892-1

Query Match 52.0%; Score 13; DB 4; Length 1365;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgc 13

Db 1139 ggcagcaggtgc 1151

RESULT 27

US-07-862-588B-6

Sequence 6, Application US/07862588B

Patent No. 5916796

GENERAL INFORMATION:

APPLICANT: Joergensen, Per Linna

APPLICANT: Sch lein, Martin

APPLICANT: Hansen, Christian

TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/862,588B

FILING DATE: 19920727

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00013

FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:

NAME: zelson, Steve T. / lambiris, Elias J.

REGISTRATION NUMBER: 30,335 / 33,728

REFERENCE/DOCKET NUMBER: 3425..204-05

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1624 base pairs

TYPE: NUCLEIC ACID

```
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match          52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcc 19
    |||
Db 934 CAGGTGCATGCC 946

RESULT 28
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnae
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB.40250
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 30..(1625.1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
; OTHER INFORMATION: (OTR)
US-07-862-588B-5

Query Match          52.0%; Score 13; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcc 19
    |||
Db 934 CAGGTGCATGCC 946

RESULT 29
US-09-258-373-21/C
; Sequence 21, Application US/09258373
; Patent No. 6150110
; GENERAL INFORMATION:
; APPLICANT: Fletcher, Jonathan A.
; APPLICANT: Xiao, Sheng
; TITLE OF INVENTION: HMG(Y)-LAMA4* FUSION ONCOGENE,
; TITLE OF INVENTION: ONCOPROTEIN AND METHODS OF USE
; FILE REFERENCE: B0801/7135/ERP
; CURRENT APPLICATION NUMBER: US/09/258,373
; CURRENT FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 60/076,401
; EARLIER FILING DATE: 1998-02-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1875
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-258-373-21

Query Match          52.0%; Score 13; DB 3; Length 1875;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ccgacgcaggtgca 14
    |||
Db 838 CGCAGCAGGTGCA 826

RESULT 30
US-09-017-706-3
; Sequence 3, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOSE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410F57
; US-09-017-706-3
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgcagcagtgca 14
        |||
Db      1530 ccgcagcagtgca 1542
```

```
RESULT 31
; US-09-017-706-4
; Sequence 4, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPENTAPOSE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410H57
; US-09-017-706-4
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgcagcagtgca 14
        |||
Db      1530 ccgcagcagtgca 1542
```

```
RESULT 32
; US-09-017-706-5
; Sequence 5, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPENTAPOSE, VECTOR CONTAINING SAID GENE AND
```

```
; TITLE OF INVENTION: TRANSFORMANT
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410OL57
; US-09-017-706-5
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgcagcagtgca 14
        |||
Db      1530 ccgcagcagtgca 1542
```

```
RESULT 33
; US-09-017-706-6
; Sequence 6, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPENTAPOSE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID: POS3410FI39
; US-09-017-706-6
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgcagcagtgca 14
        |||
Db      1530 ccgcagcagtgca 1542
```

Db 1530 cgcagcaggtgca 1542

RESULT 34

US-09-017-706-7

```

; Sequence 7, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; OTHER INFORMATION: PLASMID: POS3410H139
; US-09-017-706-7

```

Query Match 52.0%; Score 13; DB 3; Length 1981;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
 Db 1530 cgcagcaggtgca 1542

RESULT 35

US-09-017-706-8

```

; Sequence 8, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; LOCATION: (85)..(1848)

```

```

; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID: POS3410L139
; US-09-017-706-8

```

Query Match 52.0%; Score 13; DB 3; Length 1981;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
 Db 1530 cgcagcaggtgca 1542

RESULT 36

US-08-317-522A-8/C

```

; Sequence 8, Application US/08317522A
; Patent No. 5599918
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,522A
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 199..2223
; US-08-317-522A-8

```

Query Match 52.0%; Score 13; DB 1; Length 2223;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgc 13
 Db 2118 GGCAGCAGGTGC 2106

RESULT 37

US-08-439-818A-8/C

; Sequence 8, Application US/08439818A

; Patent No. 5654145

```

: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin and Trophinin-Assisting
: TITLE OF INVENTION: Proteins
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/439,818A
: FILING DATE: 12-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/317,522
: FILING DATE: 04-OCT-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 1563
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2223 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 199..2223
: US-08-439-818A-8

Query Match          52.0%; Score 13; DB 1; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggcgcagcaggtgc 13
        |||||
Db       2118 gccgcagcaggtgc 2106

RESULT 38
US-08-751-965-8/c
: Sequence 8, Application US/08751965
: Patent No. 5858360
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin and Trophinin-Assisting
: TITLE OF INVENTION: Proteins
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/751,965
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/439,818
: FILING DATE: 12-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 2252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2223 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 199..2223
: US-08-751-965-8

Query Match          52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggcgcagcaggtgc 13
        |||||
Db       2118 gccgcagcaggtgc 2106

RESULT 39
US-08-738-975-8/c
: Sequence 8, Application US/08738975
: Patent No. 5880267
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin and Trophinin-Assisting
: TITLE OF INVENTION: Proteins
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/738,975
: FILING DATE: herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/439,818
: FILING DATE: 05-Dec-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 2251
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2223 base pairs
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-738-975-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 gccgacgaggtgc 13
|||||
Db 2118 gccgacgaggtgc 2106

RESULT 40
US-08-728-626-8/c
Sequence 8, Application US/08728626
Patent No. 5910451
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 1563
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-728-626-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 gccgacgaggtgc 13
|||||

Db 2118 gccgacgaggtgc 2106

RESULT 41
US-08-808-599A-8/c
Sequence 8, Application US/08808599A
Patent No. 611089
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin, Trophinin-Assisting
PROTEINS and Methods to Inhibit Implantation
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808.599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-808-599A-8

Query Match 52.0%; Score 13; DB 3; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 gccgacgaggtgc 13
|||||
Db 2118 gccgacgaggtgc 2106

RESULT 42
US-08-584-226-1/c
Sequence 1, Application US/08584226
Patent No. 5798240
GENERAL INFORMATION:
APPLICANT: Martinis, Susan A.
APPLICANT: Sassanfar, Mandana
APPLICANT: Kim, Sunghoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-tRNA


```

; TITLE OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/584,226
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/305,766
; FILING DATE: 13-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI94-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1563
; US-08-584-226-1

Query Match          52.0%; Score 13; DB 1; Length 2290;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      7 caggtgcatgcgc 19
        |||||
Db      1233 CAGGTGCATGCC 1221

```

```

RESULT 43
US-08-941-445A-16/c
; Sequence 16, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; APPLICANT: Guan, Hanning
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941,445A

```

```

; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: translt-peptide
; LOCATION: 2..190
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: 191..2467
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..2470
; US-08-941-445A-16

```

```

Query Match          52.0%; Score 13; DB 3; Length 2763;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggcgcagcagtg 13
        |||||
Db      443 GCGCAGCAGGTGC 431

```

```

RESULT 44
US-07-872-678A-36
; Sequence 36, Application US/07872678A
; Patent No. 5541060
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme, et al.
; TITLE OF INVENTION: DETECTION OF EARLY-ONSET
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: Post Office Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,678A
; FILING DATE: 22-APRIL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: ARCD016
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3618 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-872-678A-36

Query Match 52.0%; Score 13; DB 1; Length 3618;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

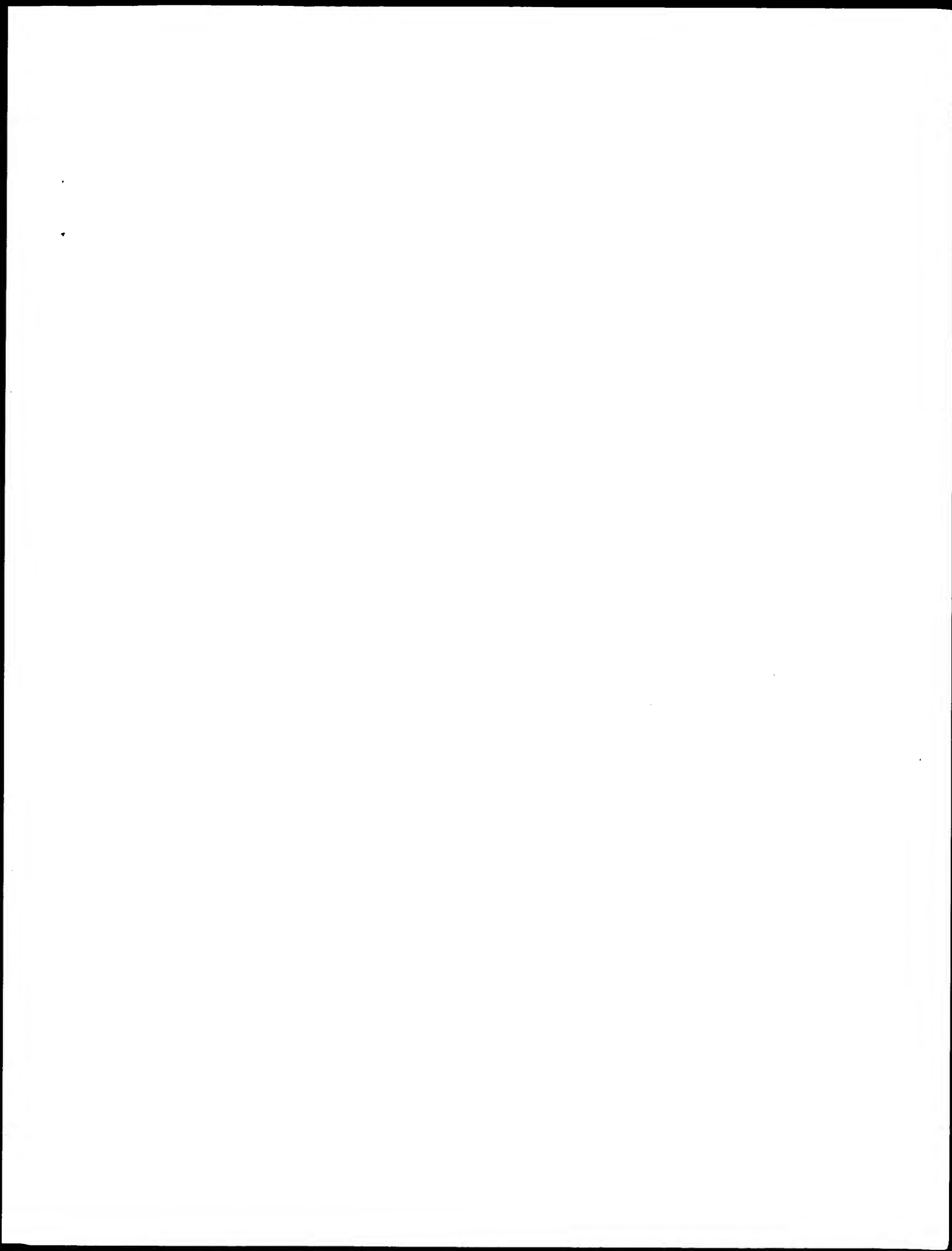
QY 2 gcgcagcaggtgca 14
|||||
Db 2446 GCGCAGCAGGTGCA 2458

RESULT 45
US-08-386-727-3/c
Sequence 3, Application US/08386727
Patent No. 5792647
GENERAL INFORMATION:
APPLICANT: ROSEMAN, SAUL
APPLICANT: BASSLER, BONNIE
APPLICANT: KEYHANI, NEMAT O.
APPLICANT: CHITLARU, EDITH
APPLICANT: ROWE, CHRIS
APPLICANT: YU, CHARLES
TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/386,727
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOBBS, ANN S.
REGISTRATION NUMBER: 36,830
REFERENCE/DOCKET NUMBER: 4130/206916
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3670 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-386-727-3

Query Match 52.0%; Score 13; DB 1; Length 3670;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgca 13
|||||
Db 2248 GCGCAGCAGGTGCA 2236

Search completed: October 9, 2001, 15:55:32
Job time: 13208 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:28 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Sequence: 1 gcgcagcagtgcatccacacatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1022815 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 520956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *
12: gb_est12: *
13: gb_est13: *
14: gb_est14: *
15: gb_est15: *
16: gb_est16: *
17: gb_est17: *
18: gb_est18: *
19: gb_est19: *
20: gb_est20: *
21: gb_est21: *
22: gb_est22: *
23: gb_est23: *
24: gb_est24: *
25: gb_est25: *
26: gb_est26: *
27: gb_est27: *
28: gb_est28: *
29: gb_est29: *
30: gb_est30: *
31: gb_est31: *
32: gb_est32: *
33: gb_est33: *
34: gb_est34: *
35: gb_est35: *
36: gb_est36: *
37: gb_est37: *
38: gb_est38: *
39: gb_est39: *
40: gb_est40: *
41: gb_est41: *
42: gb_est42: *
43: gb_est43: *

44: em_esthum10: *
45: em_esthum11: *
46: em_esthum12: *
47: em_esthum13: *
48: em_esthum14: *
49: em_esthum15: *
50: em_esthum16: *
51: em_esthum17: *
52: em_esthum18: *
53: em_esthum19: *
54: em_esthum20: *
55: em_esthum21: *
56: em_esthum22: *
57: em_esthum23: *
58: em_esthum24: *
59: em_esthum25: *
60: em_esthum26: *
61: em_esthum27: *
62: em_esthum28: *
63: em_esthum29: *
64: em_esthum30: *
65: em_esthum31: *
66: em_esthum32: *
67: em_esthum33: *
68: em_esthum34: *
69: em_esthum35: *
70: em_esthum36: *
71: em_esthum37: *
72: em_esthum38: *
73: em_esthum39: *
74: em_esthum40: *
75: em_esthum41: *
76: em_esthum42: *
77: em_esthum43: *
78: em_esthum44: *
79: em_esthum45: *
80: em_esthum46: *
81: em_esthum47: *
82: em_esthum48: *
83: em_esthum49: *
84: em_esthum50: *
85: em_esthum51: *
86: em_esthum52: *
87: em_esthum53: *
88: em_esthum54: *
89: em_esthum55: *
90: em_esthum56: *
91: em_esthum57: *
92: em_esthum58: *
93: em_esthum59: *
94: em_esthum60: *
95: em_esthum61: *
96: em_esthum62: *
97: em_esthum63: *
98: em_esthum64: *
99: em_esthum65: *
100: em_esthum66: *
101: em_esthum67: *
102: em_esthum68: *
103: em_esthum69: *
104: em_esthum70: *
105: em_esthum71: *
106: em_esthum72: *
107: em_esthum73: *
108: em_esthum74: *
109: em_esthum75: *
110: em_esthum76: *
111: em_esthum77: *
112: em_esthum78: *
113: em_esthum79: *
114: em_esthum80: *
115: em_esthum81: *
116: em_esthum82: *

```

117: qb_est48:*
118: qb_est49:*
119: qb_est50:*
120: qb_est51:*
121: qb_est52:*
122: qb_est53:*
123: qb_est54:*
124: qb_est55:*
125: qb_est56:*
126: qb_est57:*
127: qb_est58:*
128: qb_est59:*
129: qb_est60:*
130: qb_est61:*
131: qb_est62:*
132: qb_est63:*
133: qb_est64:*
134: qb_est65:*
135: qb_est66:*
136: qb_est67:*
137: qb_est68:*
138: qb_est69:*
139: qb_est70:*
140: qb_est71:*
141: qb_est72:*
142: qb_est73:*
143: qb_est74:*
144: qb_est75:*
145: qb_est76:*
146: qb_est77:*
147: qb_est78:*
148: qb_est79:*
149: qb_est80:*
150: qb_est81:*
151: qb_est82:*
152: qb_est83:*
153: qb_est84:*
154: qb_est85:*
155: qb_est86:*
156: qb_est87:*
157: qb_est88:*
158: qb_est89:*
159: qb_est90:*
160: qb_est91:*
161: qb_est92:*
162: qb_est93:*
163: qb_est94:*
164: qb_est95:*
165: qb_est96:*
166: qb_est97:*
167: qb_est98:*
168: qb_est99:*
169: qb_est100:*
170: qb_est101:*
171: qb_est102:*
172: qb_est103:*
173: qb_est104:*
174: qb_est105:*
175: qb_est106:*
176: qb_est107:*
177: qb_est108:*
178: qb_est109:*
179: qb_est110:*
180: qb_est111:*
181: qb_est112:*
182: qb_est113:*
183: qb_est114:*
184: qb_est115:*
185: qb_est116:*
186: qb_est117:*
187: qb_est118:*
188: qb_est119:*
189: qb_est120:*

```

```

190: qb_est110:*
191: qb_est111:*
192: qb_est112:*
193: qb_est113:*
194: qb_est114:*
195: qb_est115:*
196: qb_est116:*
197: qb_est117:*
198: qb_est118:*
199: qb_est119:*
200: qb_est120:*
201: qb_est121:*
202: qb_est122:*
203: qb_est123:*
204: qb_est124:*
205: qb_est125:*
206: qb_est126:*
207: qb_est127:*
208: qb_est128:*
209: qb_est129:*
210: qb_est130:*
211: qb_est131:*
212: qb_est132:*
213: qb_est133:*
214: qb_est134:*
215: qb_est135:*
216: qb_est136:*
217: qb_est137:*
218: qb_est138:*
219: qb_est139:*
220: qb_est140:*
221: qb_est141:*
222: qb_est142:*
223: qb_est143:*
224: qb_est144:*
225: qb_est145:*
226: qb_est146:*
227: qb_est147:*
228: qb_est148:*
229: qb_est149:*
230: qb_est150:*
231: qb_est151:*
232: qb_est152:*
233: qb_est153:*
234: qb_est154:*
235: qb_est155:*
236: qb_est156:*
237: qb_est157:*
238: qb_est158:*
239: qb_est159:*
240: qb_est160:*
241: qb_est161:*
242: qb_est162:*
243: qb_est163:*
244: qb_est164:*
245: qb_est165:*
246: qb_est166:*
247: qb_est167:*
248: qb_est168:*
249: qb_est169:*
250: qb_est170:*
251: qb_est171:*
252: qb_est172:*
253: qb_est173:*
254: qb_est174:*
255: qb_est175:*
256: qb_est176:*
257: qb_est177:*
258: qb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	72.0	456	234	A0852799	A0852799 LMAJFV1_1
2	18	72.0	376	171	B9333544	B9333544 CM1-NT027
3	16	64.0	399	156	D24778	D24778 RLCR2545A_R
4	16	64.0	479	234	A0852582	A0852582 LMAJFV1_1
5	16	64.0	439	258	P968R	A1390674 Leishmani
6	16	64.0	761	170	B9864751	B9864751 963054F05
7	16	64.0	915	150	B9581354	B9581354 602100740
8	15	60.0	164	170	B9836559	B9836559 CM2-HT096
9	15	60.0	226	234	A0851231	A0851231 LMAJFV1_1
10	15	60.0	266	19	A1410504	A1410504 EST238797
11	15	60.0	352	10	A6680915	A6680915 EFMFM050
12	15	60.0	380	12	AA800788	AA800788 EST190285
13	15	60.0	396	167	BE429960	BE429960 MTD013_H0
14	15	60.0	418	9	AA625125	AA625125 at70a04.r
15	15	60.0	453	108	A0166733	A0166733 A0166733
16	15	60.0	463	167	BE466670	BE466670 WHEI139_F
17	15	60.0	475	12	AA821080	AA821080 GMD09622.5
18	15	60.0	475	173	BG101817	BG101817 RHIZ22_22-
19	15	60.0	487	10	AA680808	AA680808 HI03757.5-
20	15	60.0	499	171	BF916376	BF916376 IL3-UT011
21	15	60.0	505	156	C98143	C98143 C98143 Rice
22	15	60.0	514	154	BG467085	BG467085 IAO4C06_B
23	15	60.0	524	104	AI957283	AI957283 u185e12.x
24	15	60.0	536	174	BG32022	BG32022 naf35e06.
25	15	60.0	545	237	A2071920	A2071920 RPT-23-3
26	15	60.0	553	241	A2283519	A2283519 RPT-23-1
27	15	60.0	567	156	C98142	C98142 C98142 Rice
28	15	60.0	575	149	BF489544	BF489544 AT25594.5
29	15	60.0	605	233	A0783596	A0783596 HS_3121.A
30	15	60.0	623	173	BG103212	BG103212 RHIZ22_19
31	15	60.0	730	154	BG479445	BG479445 602525920
32	15	60.0	761	164	BE194330	BE194330 HVSMBH008
33	15	60.0	776	149	BF493796	BF493796 AT02150.8
34	15	60.0	792	152	BG299490	BG299490 HVSMBH002
35	15	60.0	824	155	BG309144	BG309144 HVSMBH000
36	15	60.0	843	104	AJ282984	AJ282984 4A3A-P285
37	15	60.0	843	139	BE731371	BE731371 601565029
38	15	60.0	886	168	BE684687	BE684687 602141056
39	14	56.0	116	139	BE709927	BE709927 IL3-HT061
40	14	56.0	119	170	BF876173	BF876173 CM4-ET009
41	14	56.0	178	246	AZ595033	AZ595033 LM0407F22
42	14	56.0	196	143	BF017303	BF017303 uW34H08.x
43	14	56.0	206	24	A1714341	A1714341 UT-R-AE1-
44	14	56.0	219	161	BB563629	BB563629 BB563629
45	14	56.0	226	114	AW326327	AW326327 18754 MAR

ALIGNMENTS

TITLE	Blistrain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko J., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M., Harvey,N., McCann,R., Tsagarelshvili,R., Williams,T., Jackson X., Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M. A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA
JOURNAL	Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
COMMENT	Contact: Akopyants, NS / Beverley, SM WashU Leishmania Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu Library construction: Natalia S. Akopyants, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center If using this information please cite: N.S. Akopyants and S.M. Beverley 'A survey of the leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley ph.D. (beverley@borcim.wustl.edu) Seq primer: -40RP from Gibco Class: shotgun High quality sequence stop: 252.
FEATURES	Location/Qualifiers
Source	1..456 /organism="Leishmania major" /strain="Friedlin strain VI" /db_xref="taxon:5664" /clone="LMAjPVI_lm77g909" /clone_lib="Leishmania major FV1 random genomic library" /lab_host="TOP10 (Invitrogen)" /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV; genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNase polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT	78 a 141 c 144 g 93 t
ORIGIN	
Query Match	72.0%; Score 18; DB 234; Length 456;
Best Local Similarity	100.0%; Pred. No. 4.1;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
QY	1 gcgcacagctcatcgc 18
Db	253 GCGCACAGCTCATCGC 270

FEATURES

Source

ORIGIN

Query Mat

st Local

Qy 1 9

ALIGNMENTS

TITLE	Blistrain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko J., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M., Harvey,N., McCann,R., Tsagarelshvili,R., Williams,T., Jackson X., Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M. A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA
JOURNAL	Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
COMMENT	Contact: Akopyants, NS / Beverley, SM WashU Leishmania Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu Library construction: Natalia S. Akopyants, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center If using this information please cite: N.S. Akopyants and S.M. Beverley 'A survey of the leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley ph.D. (beverley@borcim.wustl.edu) Seq primer: -40RP from Gibco Class: shotgun High quality sequence stop: 252.
FEATURES	Location/Qualifiers
Source	1..456 /organism="Leishmania major" /strain="Friedlin strain VI" /db_xref="taxon:5664" /clone="LMAJFVI_lm77g909" /clone_lib="Leishmania major FV1 random genomic library" /lab_host="TOP10 (Invitrogen)" /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV; genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNase polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT	78 a 141 c 144 g 93 t
ORIGIN	
Query Match	72.0%; Score 18; DB 234; Length 456;
Best Local Similarity	100.0%; Pred. No. 4.1;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
QY	1 gcgcacagctcatcgc 18
Db	253 GCGCACAGCTCATCGC 270

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
200202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lcr-CM1<2-CM1-NT0270-
211200-661-g03<3=2000-12-21<4=1)

Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.

FEATURES
source

1..376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NT0270"
/dev_stage="Adult"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT
ORIGIN
77 a 104 c 99 g 96 t

Query Match
Best Local Similarity 100.0%; Score 16; DB 171; Length 376;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggtgcatcgccagcat 24
|||||
Db 329 ggtgcatcgccagcat 314

RESULT 3
D24778 399 bp mRNA EST 08-JUL-1999
LOCUS R1CR2545A Rice root Oryza sativa cDNA clone R2545_1A, mRNA
DEFINITION sequence.
ACCESSION D24778
VERSION D24778.1 GI:428626
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 399)
Minobe,Y. and Sasaki,T.
Rice cDNA from root
Unpublished (1995)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
PRODUCT "RCP".

FEATURES
source
Location/Qualifiers
1..399
/organism="Oryza sativa"

/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R2545_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."

BASE COUNT 114 a 70 c 102 g 112 t 1 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 16; DB 156; Length 399;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggtgcatcgccagcat 24
|||||
Db 251 ggtgcatcgccagcat 266

RESULT 4
A0852582 479 bp DNA GSS 13-APR-2001
LOCUS LMAJFV1_lm74c04.y1 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1_lm74c04 5', DNA sequence.
ACCESSION A0852582
VERSION A0852582.1 GI:6118907
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 479)
Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
Kissinger,J., Roos,D.S., Maria,M., Hillier,L., Chinwalla,A.,
Blistain,A., Schmitt,A., Person,B., Theising,B., Ritzer,E., Ronko,
I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
Harvey,N., McCann,R., Tsagarisvilli,R., Williams,T., Jackson,Y.,
Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
A survey of the Leishmania major Friedlin strain VI genome by
shotgun sequencing: a resource for DNA
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@bpcrcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@bpcrcim.wustl.edu)
Seq primer: -408P from Gibco
Class: shotgun
High quality sequence stop: 372.

FEATURES
source

1..479
Location/Qualifiers
/organism="Leishmania major"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm74c04"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 98 a 155 c 147 g 79 t
 ORIGIN

Query Match 64.0%; Score 16; DB 234; Length 479;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cagcaggtcgcgc 19
 |||
 Db 47 CAGCAGTGCATGCC 62

RESULT 5
 P968R 548 bp DNA GSS 25-JUL-2000
 LOCUS Leishmania major Friedlin PAC P968 right end-sequence, genomic
 DEFINITION survey sequence.
 ACCESSION AL390674 GI:9501650
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania
 1 (bases 1 to 548)
 Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
 Smith,D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 548)
 Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
 Barrell,B.G.
 DIRECT SUBMISSION
 TITLE Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing
 JOURNAL project, Sanger Centre, The Wellcome Trust Genome Camps, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 allcats@sanger.ac.uk
 COMMENT see <http://www.ebi.ac.uk/parasites/leish.html>
 details of Leishmania sequencing at the Sanger Centre are available
 at <http://www.sanger.ac.uk/Projects/L.major/>
 The primer sequence can be obtained from allcats.sanger.ac.uk.

FEATURES
 Location/Qualifiers
 1..348
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="PAC P968"

BASE COUNT 86 a 167 c 163 g 132 t
 ORIGIN

Query Match 64.0%; Score 16; DB 258; Length 548;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcagtcgcagcatt 25
 |||
 Db 541 GTGCATGCCAGCAT 526

RESULT 6
 BF864751 761 bp mRNA EST 19-JAN-2001
 LOCUS BF864751
 DEFINITION 963054E05.Y1 C. reinhardtii CC-1690, stress condition I, normalized
 Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BF864751 GI:12254895
 VERSION BF864751.1 GI:12254895
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE
 AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
 Lefebvre,P., McDermott,J.P., Strager,J., Sillow,C. and Stern,D.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 JOURNAL Unclutellar System for Analyzing Gene Function and Regulation in
 Vascular Plants: project phase 3
 COMMENT Unpublished (2000)
 CONTACT: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu

FEATURES
 source
 Location/Qualifiers
 1..761
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, stress condition I,
 normalized, Lambda zap II"
 /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into Lambda zap II
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 Bluescript II SK- plasmids were excised from the Lambda
 zap clones by superinfection with Exassist (Stratagene)
 phage. The library was normalized using method 4 described
 in Honalido et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 179 c 308 g 117 t
 ORIGIN

Query Match 64.0%; Score 16; DB 170; Length 761;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcagtcgcagca 23
 |||
 Db 466 AGGTGCATGCCAGCA 451

RESULT 7
 BF581354 915 bp mRNA EST 12-DEC-2000
 LOCUS BF581354
 DEFINITION 602100740F1 NCI CGAP_C024 Mus musculus cDNA clone IMAGE:4224225 5',
 mRNA sequence.
 ACCESSION BF581354
 VERSION BF581354.1 GI:11655066
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 915)
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM9813 row: 0 column: 10
High quality sequence stop: 646.
Location/Qualifiers
1. .915

FEATURES

source

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4224225"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SpORt; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb, constructed by LITE
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT

218 a 259 c 230 g 208 t

ORIGIN

Query Match 64.0%; Score 16; DB 150; Length 915;
Best Local Similarity 100.0%; Pred. NO. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 accaggtcagtcacca 20
|||||

Db 737 AGCAGTCATCCCA 752

RESULT 8
LOCUS BF836559 164 bp mRNA EST 13-JAN-2001
DEFINITION CM2-HT0968-181100-508-e11 HT0968 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF836559
VERSION BF836559.1 GI:12188510
KEYWORDS EST.

SOURCE

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
/M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE 2020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=CM2&t2=CM2-HT0968-
181100-508-e11&t3=2000-11-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 95
High quality sequence stop: 164.
Location/Qualifiers
1. .164

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0968"
/dev_stage="Adult"
/note="Organ: head-neck; Vector: puc18; Site: 1: SmaI;
Site: 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application

BASE COUNT

No. 196/716 - Ludwig Institute for Cancer Research)
Profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

46 a 39 c 45 g 34 t

ORIGIN

Query Match 60.0%; Score 15; DB 170; Length 164;
Best Local Similarity 100.0%; Pred. NO. 1.8e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccagcaggtcgtcat 15
|||||

Db 28 GCCCAGCAGTCGAT 14

RESULT 9

AQ851231 226 bp DNA GSS 13-APR-2001
LOCUS LMAJFV1_lm44h03.x1 Leishmania major FV1 random genomic library

DEFINITION Leishmania major genomic clone LMAJFV1_lm44h03 3', DNA sequence.
ACCESSION AQ851231
VERSION AQ851231.1 GI:6055879
KEYWORDS GSS.

SOURCE Leishmania major.
Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1 (bases 1 to 226)
Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A.,
Blistain, A., Schmitt, A., Person, B., Theising, B., Riteir, E., Ronko,
I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
Harvey, N., McCann, R., Tsagaris, V., Williams, T., Jackson, Y.,
Bowers, Y., Swaller, T., Waterson, R., Wilson, R. and Beverley, S.M.
A survey of the Leishmania major Friedlin strain V1 genome by
shotgun sequencing: a resource for DNA

JOURNAL COMMENT
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Other-GSS: lm44h03.y1
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40UP from Glibco
Class: shotgun

High quality sequence stop: 226.
Location/Qualifiers
1. .226

FEATURES

source

/organism="Leishmania major"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm44h03"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site: 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments. blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 40 a 67 c 64 g 55 t
 ORIGIN

Query Match 60.0%; Score 15; DB 234; Length 226;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ccgacgaggtgacgc 16
 |||||||
 Db 88 CGCAGCAGTGCATC 102

RESULT 10
 A1410504 266 bp mRNA EST 09-FEB-1999
 LOCUS
 DEFINITION EST238797 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
 RHECX29 3' end, mRNA sequence.
 A1410504
 ACCESSION A1410504.1 GI:4254008
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus sp.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 266)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 source
 1..266
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RHECX29"
 /clone_lib="Normalized rat heart, Bento Soares"
 /note="Organ: heart; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 56 a 61 c 64 g 85 t
 ORIGIN

Query Match 60.0%; Score 15; DB 19; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatgcagca 23
 |||||||
 Db 90 GGTGATCGCAGCA 104

RESULT 11
 AA680915 352 bp mRNA EST 10-DEC-1998
 LOCUS
 DEFINITION LmFRA0507 Leishmania major Amastigote full length cDNA library
 Leishmania major cDNA clone Q83 5', mRNA sequence.
 AA680915
 ACCESSION AA680915.1 GI:3886275
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Leishmania major.
 Leishmania major
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 352)
 BASE COUNT 71 a 100 c 106 g 103 t

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Morrish,A.R., Dyall,S.D., Smith,D.F. and Blackwell,J.M.
 Analysis of Leishmania Major Amastigote Expressed Sequence Tags
 Unpublished (1997)
 On Dec 5, 1997 this sequence version replaced gi:2662920.
 Contact: Blackwell JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk
 PCR Primers
 FORWARD: GTAAACGACGCGCAGT
 BACKWARD: GGAACAGCTATGACATG
 Seq primer: AATTAACTCTCCTAAGGG
 High quality sequence stop: 352.
 Location/Qualifiers
 source
 1..352
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="Q83"
 /clone_lib="Leishmania major Amastigote full length cDNA
 library"
 BASE COUNT 70 a 110 c 86 g 86 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 cagcaggtgacgc 18
 |||||||
 Db 183 CAGCAGTGCATCGC 169

RESULT 12
 AA800788 380 bp mRNA EST 30-APR-1998
 LOCUS
 DEFINITION EST190285 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
 RU0AL45 3' end, mRNA sequence.
 AA800788
 ACCESSION AA800788.1 GI:2863743
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus sp.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 380)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 source
 1..380
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2008206"
 /db_xref="taxon:10118"
 /clone="RU0AL45"
 /clone_lib="Normalized rat lung, Bento Soares"
 /note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 71 a 100 c 106 g 103 t

ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggtgcatgcacgca 23
 |||
 Db 41 GGTGATGCCAGCA 55

RESULT 13
 BE429060/c 396 bp mRNA EST 26-JUL-2000
 LOCUS MTD013.H06F990621 ITEC MTD Durum Wheat Root Library Triticum
 DEFINITION turgidum subsp. durum cDNA clone MTD013.H06, mRNA sequence.
 ACCESSION BE429060
 VERSION BE429060.1 GI:9426903
 KEYWORDS EST.
 SOURCE durum wheat.
 ORGANISM Triticum turgidum subsp. durum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 ; Triticeae; Triticum.
 1 (bases 1 to 396)
 REFERENCE Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Hermann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
 Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M., and Wenzel,G.
 International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 CONTACT: Joudrier P
 INRA, Unite de Biochimie et Biologie Moleculaire des Cereales
 2, place VIALA, 34060 Montpellier cedex 01 FRANCE
 Tel: 33 4 99 61 23 84
 Fax: 33 4 99 61 23 48
 Email: joudrier@enscm.inra.fr
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 FEATURES
 Location/Qualifiers
 1..396
 /organism="Triticum turgidum subsp. durum"
 /cultivar="Silliana"
 /db_xref="taxon:4567"
 /clone="MTD013.H06"
 /clone_lib="ITEC MTD Durum Wheat Root Library"
 /issue_type="root"
 /dev_stage="3-day-old seedling, water-stressed"
 /note="Vector: pSPORT1; T7 primers used. See pSPORT1
 polylinker site 0.3-2.0 kbp average insert size."
 BASE COUNT 96 a 87 c 116 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 396;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagtcacgcgca 20
 |||
 Db 75 GCAGTGCGCATCGCA 61

RESULT 14
 AA625125 418 bp mRNA EST 02-MAR-1998
 LOCUS af70a04.r1 Soares.NHMPu_S1 Homo sapiens cDNA clone IMAGE:1047342
 DEFINITION 5', mRNA sequence.
 ACCESSION AA625125

VERSION AA625125.1 GI:2537510
 EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 418)
 REFERENCE Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 J., Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)
 CONTACT: Wilson RK
 JOURNAL
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 626 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 408.
 FEATURES
 Location/Qualifiers
 1..418
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1047342"
 /clone_lib="Soares.NHMPu_S1"
 /issue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site 1: Not I;
 Site 2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NBH, pregnant uterus
 NBHPU, and fetal heart NBH19W) were mixed and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 BASE COUNT 128 a 93 c 114 g 83 t

ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgacagtgatc 16
 |||
 Db 399 GCACAGGATGATC 413

RESULT 15
 AUI66733/c 453 bp mRNA EST 23-JAN-2001
 LOCUS AUI66733 Rice callus Oryza sativa subsp. japonica cDNA clone C50066
 DEFINITION AUI66733 Rice callus Oryza sativa subsp. japonica cDNA clone C50066
 , mRNA sequence.
 ACCESSION AUI66733
 VERSION AUI66733.1 GI:12405132
 KEYWORDS EST.
 SOURCE Oryza sativa subsp. japonica.
 ORGANISM Oryza sativa subsp. japonica
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatoidae; Oryzae; Oryza.
 1 (bases 1 to 453)
 REFERENCE Sasaki,T. and Yamamoto,K.
 AUI66733
 AUTHORS Rice cDNA from callus (2001)
 TITLE

JOURNAL Unpublished (2001)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abrr.affrc.go.jp
PROJECT = "RGP".
C50066_982.

FEATURES
source Location/Qualifiers
1. 453
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:39947"
/clone_lib="C50066"
/clone_lib="Rice callus"
/note="Vector: Bluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of Bluescript II SK+ phagemid."

BASE COUNT 164 a 83 c 140 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 108; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgacatt 25
|||||
DB 349 TGCAATCCGACGACATT 335

RESULT 16
BE446670 463 bp mRNA EST 25-JUL-2000
LOCUS WHE1139_F07_L1325 Wheat etiolated seedling root normalized cDNA
DEFINITION library Triticum aestivum cDNA clone WHE1139_F07_L13, mRNA
sequence.
ACCESSION BE446670
VERSION BE446670.1 GI:9446232
KEYWORDS EST.
SOURCE Triticum aestivum
ORGANISM bread wheat.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.
1 (bases 1 to 463)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,
Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
The structure and function of the expressed portion of the wheat
genomes. Normalized root cDNA library
Unpublished (2000)

JOURNAL COMMENT
CONTACT: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.
Location/Qualifiers

FEATURES
source 1. 463
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1139_F07_L13"

BASE COUNT 119 a 98 c 132 g 114 t
ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 463;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcaagtgatcgaccca 20
|||||
DB 111 GCAGGTGATCGACCA 97

RESULT 17
AA821080 475 bp mRNA EST 25-NOV-1998
LOCUS GM09622.5prime GM Drosophila melanogaster ovary Bluescript
DEFINITION Drosophila melanogaster cDNA clone GM09622.5prime, mRNA sequence.
ACCESSION AA821080
VERSION AA821080.1 GI:2890948
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 475)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)

JOURNAL COMMENT
CONTACT: Harvey D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 96 row: B column: 10
High quality sequence stop: 439.
Location/Qualifiers

FEATURES
source 1. 475
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GM09622"
/clone_lib="GM Drosophila melanogaster ovary Bluescript"
/dev_stage="newly eclosed females; germination-stage 6"
/sex="female"
/lab_host="SOLR"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Constructed using StrataGene Zap-cDNA
Synthesis kit. Oligo dT-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 128 a 147 c 121 g 79 t
 ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 475;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 agcagtcgcatgcgc 22
 |||
 Db 344 AGCGTCGTCGTCGC 358

RESULT 18
 BG101817 475 bp mRNA EST 30-JAN-2001
 LOCUS RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum proproinquum cDNA, mRNA
 DEFINITION
 ACCESSION BG101817 GI:12616762
 VERSION BG101817
 KEYWORDS EST.
 SOURCE Sorghum proproinquum.
 ORGANISM Sorghum proproinquum.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 475)
 Cordonier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
 L.H.
 TITLE An EST database from Sorghum: Sorghum proproinquum rhizomes
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@prattuga.edu
 Seq primer: PolyTMix
 High quality sequence start: 11
 High quality sequence stop: 408
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..475
 /organism="Sorghum proproinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH122)"
 /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
 Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
 from poly-A RNA in the cloning vector Lambda Zap II.
 Clones to be sequenced by mass excision."
 BASE COUNT 128 a 107 c 113 g 127 t
 ORIGIN

Query Match 60.0%; Score 15; DB 173; Length 475;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagtcgcatgcgc 20
 |||
 Db 240 GCAGTCGTCGTCGC 254

RESULT 19
 AA698088 487 bp mRNA EST 27-NOV-1998
 LOCUS HL03757.5prtime HL Drosophila melanogaster head Bluescript
 DEFINITION
 ACCESSION AA698088
 VERSION AA698088.1 GI:2701017
 KEYWORDS EST.
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 487)
 Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
 P., Lewis, S. and Rubin, G.M.
 BDGP/HMI Drosophila EST Project
 Unpublished (1997)
 TITLE JOURNAL
 COMMENT Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 USA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 Plate: 37 row: E column: 9
 High quality sequence stop: 331.

FEATURES
 source Location/Qualifiers
 1..487
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="HL03757"
 /clone_lib="HL Drosophila melanogaster head Bluescript"
 /sex="male and female"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Organ: head-brain & sensory organ; Vector:
 Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
 using Stratagene Zap-cDNA synthesis kit. Oligo dT-primed
 and directionally cloned at EcoRI and XhoI in Bluescript
 SK(+/-)."
 BASE COUNT 123 a 151 c 128 g 85 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 487;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 agcagtcgcatgcgc 19
 |||
 Db 109 AGCAGTCGTCGTCGC 95

RESULT 20
 BF916376 499 bp mRNA EST 18-JAN-2001
 LOCUS IL3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.
 DEFINITION
 ACCESSION BF916376
 VERSION BF916376.1 GI:12307834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 499)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Megal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 TITLE JOURNAL
 MEDLINE
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?PI=IL3&L2=IL3-UT0115-111200-377-F05&L3=2000-12-11&L4=1)

Seq primer: puc 18 forward
High quality sequence stop: 496.

FEATURES

Source

1. 499

/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone_lib="UT0115"
/dev_stage="Adult"

/note="Organ: uterus; tumor; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 121 a 121 c 114 g 143 t
ORIGIN

Query Match 60.0%; Score 15; DB 171; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcaagtcacgcgca 20
|||||
Db 216 GCAGTCGATCGCCA 230

RESULT 21

C98143/c

LOCUS C98143 505 bp mRNA EST 29-APR-1999
DEFINITION C98143 Rice callus Oryza sativa subsp. japonica cDNA clone C0777_7A

/mRNA sequence.

ACCESSION C98143 D28202
VERSION C98143.1 GI:3760889

KEYWORDS

EST.

SOURCE

Oryza sativa subsp. japonica.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

AUTHORS

1 (bases 1 to 505)
Yamamoto, K. and Sasaki, T.

JOURNAL

Rice cDNA from callus (1998)

COMMENT

On Oct 19, 1998 this sequence version replaced gi:454479.

CONTACT

Takuji Sasaki

NATIONAL

National Institute of Agrobiological Resources

RICE

Genome Research Program

2-1-2

Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@nri.affrc.go.jp

D28202; Submitted (01-JAN-1994)

PROJECT

"RGP"

POLYA-NO.

Location/Qualifiers

1. 505

/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"

/db_xref="taxon:39947"
/clone="C0777_7A"

/clone_lib="Rice callus"
/note="Vector: pluscript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site

BASE COUNT 158 a 103 c 170 g 72 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgcatt 25
|||||
Db 475 TGCATCGCCAGCAT 461

RESULT 22

BG467085

LOCUS BG467085 514 bp mRNA EST 20-MAR-2001
DEFINITION BG467085 Bovine Mixed Adipose cDNA library Bos taurus cDNA 5', mRNA

sequence.

ACCESSION BG467085.1 GI:13396060
VERSION BG467085.1
KEYWORDS EST.

SOURCE

cow.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 514)

AUTHORS

Moore, S.S., Hansen, C., Li, C., Fu, A., Meng, Y., Li, G., Murdoch, G.,
Dixon, W., and Christoperson, B.

TITLE

cDNA's from bovine subcutaneous adipose tissue

JOURNAL

Unpublished (2001)

COMMENT

Contact: Dr. Stephen Moore

DEPT

Dept of AFNS, University of Alberta

TEL

410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

FAX

780 492 4265

EMAIL

smoore@afns.ualberta.ca

NOTE

The sequence best matches gb:FNMDH08 (B. taurus C1-B14.5b mRNA for
NADH dehydrogenase (ubiquinone)) in main database at high score of
908.0 and E-value of 0.0

PCR

PRIMERS

FORWARD

M13 Forward

REVERSE

M13 Reverse

SEQ

Primer: T3 primer

HIGH

quality sequence stop: 514

POLYA-NO.

Location/Qualifiers

1. 514

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="Bovine Mixed Adipose cDNA library"

/sex="two males and one female mixed"

/tissue_type="Adipocyte"

/cell_type="Adipocyte"

/dev_stage="young adult"

/lab_host="XLI-BlueMRF'scrain"

/note="Organ: Subcutaneous, Omental, Mesenteric adipose;
Vector: Uni-ZAPXR; Site_1: EcoRI; Site_2: Xho I"

BASE COUNT 115 a 112 c 132 g 155 t

ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 514;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgcatt 25

Db 242 TGCATCGCCAGCAT 256

```

RESULT 23
LOCUS   A1957283 524 bp mRNA EST 20-AUG-1999
DEFINITION U185e12.x1 Sugano mouse kidney mklia Mus musculus cDNA clone
IMAGE:2158894 3', mRNA sequence.
ACCESSION A1957283
VERSION   A1957283.1 GI:5749992
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 524)
AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterslon,R. and Wilson,R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
TITLE     JOURNAL
COMMENT   Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1002066
Seq primer: custom primer used
High quality sequence stop: 359.
Location/Qualifiers
1..524
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:2158894"
/clone_lib="Sugano mouse kidney mklia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pMR18-FL3; Site_1: DraIII
(CACCTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pMR18-FL3
vector (5' site CACCTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGTCTCTTAAGAAGCTCGG and 3' end
primer CGACCTGACGCTCGAGACA."
BASE COUNT 151 a 102 c 147 g 124 t
ORIGIN
Query Match 60.0%; Score 15; DB 104; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 tgcattgcacgcatt 25
DB 352 TGCATGCGCCGACAT 338
RESULT 24
LOCUS   BG232022 536 bp mRNA EST 09-FEB-2001
DEFINITION BG232022.x1 Soares_NPBMC Homo sapiens cDNA clone IMAGE:4143059 3'
similar to TR:P79434 P79434 CONNECTING PEPTIDE/TRANSMEMBRANE
/cytoplasmic DOMAIN;; mRNA sequence.

```

```

ACCESSION BG232022
VERSION   BG232022.1 GI:12727177
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS   NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail@nih.gov
Tissue Procurement: M. Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D. and M. Fatima
Bonaldo, Ph.D.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 412.
Location/Qualifiers
1..536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4143059"
/clone_lib="Soares_NPBMC"
/tissue_type="lymphocyte"
/lab_host="DH10B (phage-resistant)"
/note="Organ: blood; Vector: p773D-Pac; Site_1: NotI;
Site_2: EcoRI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCGCTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacial), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. Library
is normalized; constructed in the laboratory of M. Bento
Soares (University of Iowa)."
BASE COUNT 152 a 119 c 133 g 132 t
ORIGIN
Query Match 60.0%; Score 15; DB 174; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 gcaagtgcatcgcca 20
DB 293 GCAGGTGATCGCCA 279
RESULT 25
LOCUS   AZ071920 545 bp DNA GSS 31-MAR-2000
DEFINITION RPCI-23-395D19.TUB RPCI-23 Mus musculus genomic clone
ACCESSION RPCI-23-395D19, DNA sequence.
VERSION   AZ071920
KEYWORDS  AZ071920.1 GI:7364817
SOURCE    GSS.
ORGANISM  house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 545)
AUTHORS   Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akioret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
TITLE     JOURNAL
COMMENT   Other_GSSs: RPCI-23-395D19.TV

```


Contact: Shayang Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://resgen.com>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/Bac_end_intro.html
Plate: 395 row: D column: 19
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..545

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-395D19"

/clone_1ib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

137 a 153 c 121 g 134 t

ORIGIN

Query Match 60.0%; Score 15; DB 237; Length 545;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagtgatcgccca 20

Db 505 GCAGTGCATCGCCA 491

RESULT 26
A2283519 553 bp DNA GSS 27-JUL-2000
LOCUS RPCI-23-125E16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-125E16

DEFINITION
' DNA sequence.
ACCESSION
A2283519
VERSION
A2283519.1 GI:9525226
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus.

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 553)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Aklioret
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, F.
and Fraser, C. M.
Mouse BAC End Sequences from Library RPCI-23

TITLE
JOURNAL
COMMENT
Unpublished (1999)
Other_GSSs: RPCI-23-125E16.TV
Contact: Shayang Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong

FEATURES

source

Location/Qualifiers

1..553

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-125E16"

/clone_1ib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

135 a 156 c 126 g 136 t

ORIGIN

Query Match 60.0%; Score 15; DB 241; Length 553;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagtgatcgccca 20

Db 525 GCAGTGCATCGCCA 511

RESULT 27

C98142 567 bp mRNA EST 19-OCT-1998
LOCUS C98142 Rice callus Oryza sativa subsp. japonica cDNA clone

DEFINITION
C0777_102, mRNA sequence.
ACCESSION
C98142
VERSION
C98142.1 GI:3760888
KEYWORDS
EST.

ORGANISM
Oryza sativa subsp. japonica.
Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE
AUTHORS
Ehhardtideae; Oryzae; Oryza.
1 (bases 1 to 567)
Sasaki, T. and Minobe, Y.
Rice cDNA from callus
Unpublished (1994)

JOURNAL

COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
PROJECT = "RGP".

FEATURES

source

Location/Qualifiers

1..567

/organism="Oryza sativa subsp. japonica"

/strain="cultivar Nipponbare, sub-species japonica"

/db_xref="taxon:39947"

/clone="C0777_102"

/clone_1ib="Rice callus"

/note="Vector: pBluescript II SK+; Site: 1; SalI; Site: 2; NotI; cDNA prepared from rice callus RNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site

BASE COUNT 190 a 105 c 196 g 75 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 567;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacacatc 25
|||||
Db 472 TGCATCCGACGACAT 458

RESULT 28
BF489544/c 575 bp mRNA EST 16-APR-2001
LOCUS
DEFINITION AT255594.Sprime At Drosophila melanogaster adult testes pOTB7
ACCESSION BF489544
VERSION BF489544.1 GI:11572845
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 575)
Stapleton, M., Broksstein, P., Hong, L., Agbayan, A., Baxter, E., Berman, B., Carlson, J., Champagne, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nuno, J., Pacled, J., Park, S., Paragas, V., Phouanavong, S., Man, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
unpublished (2000)
CONTACT: Stapleton, M.
BDGP

TITLE
JOURNAL
COMMENT
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003597
Plate: AT.255 row: H column: 10
High quality sequence stop: 536.
Location/Qualifiers
1..575
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT25594"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates AT.121-AT.319: DHS-alpha Tona"
/note="Organ: ADULT testes; Vector: pOTB7; Site: 1: RCOBT; Site 2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 114 a 167 c 162 g 132 t
ORIGIN

Query Match 60.0%; Score 15; DB 149; Length 575;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgacatc 16
|||||
Db 520 CGCAGCAGGTGACATC 506

RESULT 29
AO783596/c 605 bp DNA GSS 03-AUG-1999
LOCUS
DEFINITION HS_3121_A2_F07_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3121 COL=14 Row=K, DNA sequence.
ACCESSION AO783596
VERSION AO783596.1 GI:5691150
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 605)
Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D., and Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (inforesgen.com).
BAC end Web Server: <http://www.husc.washington.edu>
Plate: 3121 row: K column: 14
Seq primer: 17
Class: BAC ends
High quality sequence stop: 605.
Location/Qualifiers
1..605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=3121 COL=14 Row=K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 187 a 187 c 123 g 92 t 16 others
ORIGIN

Query Match 60.0%; Score 15; DB 233; Length 605;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtcatgcacacatc 24
|||||
Db 248 GTGCATCCGACGACAT 234

RESULT 30
BG103212 623 bp mRNA EST 30-JAN-2001
LOCUS
DEFINITION RH122_19_E10_g1_A003 Rhizome2 (RH122) Sorghum prolinquum cDNA, mRNA sequence.
ACCESSION BG103212
VERSION BG103212.1 GI:12618045
KEYWORDS EST.
SOURCE Sorghum prolinquum.
ORGANISM Sorghum prolinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 623)
Cordonnier-Pratt, M.-M., Gingle, A., Peterson, A., Sudman, M. and Pratt, L. H.
An EST database from Sorghum: Sorghum prolinquum rhizomes

REFERENCE
AUTHORS
TITLE

JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolyT/Mix
High quality sequence start: 48
High quality sequence stop: 560
POLYA-No.

FEATURES Location/Qualifiers
1..623
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RH12)"
/note="Organ: Rhizomes; Vector: pBluescript II from lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 171 a 140 c 149 g 163 t
ORIGIN

Query Match 60.0%; Score 15; DB 173; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcagatgcacgcga 20
|||||
Db 394 GCAGGTGATCGCA 408

RESULT 31
LOCUS BE479445 730 bp mRNA EST 21-MAR-2001
DEFINITION 602525920F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649525 5',
mRNA sequence.
ACCESSION BE479445
VERSION BE479445.1 GI:13411724
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 730)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1429 row: h column: 06
High quality sequence stop: 642.

FEATURES Location/Qualifiers
1..730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4649525"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAC(g). Size-selected >500bp

for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 169 a 184 c 193 g 184 t
ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 730;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcagatgcac 16
|||||
Db 706 CGCAGCAGGTGATC 692

RESULT 32
LOCUS BE194330 761 bp mRNA EST 02-MAR-2001
DEFINITION HVSMEH0085C19f Hordeum vulgare 5-45 DAP spike EST library
HVCNMA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0085C19f,
mRNA sequence.
ACCESSION BE194330
VERSION BE194330.2 GI:13187270
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
1 (bases 1 to 761)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Fritsch, D., Yu
Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: AATTAACTCTCACTAAGG
High quality sequence stop: 647.

FEATURES Location/Qualifiers
1..761
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEH0085C19f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library
HVCNMA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 198 a 150 c 201 g 212 t
ORIGIN

Query Match 60.0%; Score 15; DB 164; Length 761;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtgcacgcacacat 24
|||||
Db 54 GTGCATCGCACACAT 40

RESULT 33
BF493796/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF493796 776 bp mRNA EST 16-APR-2001
AT02150.Sprime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT02150 5 similar to CG9085:
BF493796
BF493796.1 GI:11577097
EST.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 776)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nuno, J., Paclob, J.,
Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E.,
Celniker, S. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
COMMENT
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic sequence AE003597
Plate: AT.21 row: E column: 2
High quality sequence stop: 670.
Location/Qualifiers
1..776
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT02150"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT 10-AT 120: DH5-alpha. Plates
AT 121-AT 319: DH5-alpha Tona"
/note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcORI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

BASE COUNT 168 a 205 c 211 g 192 t

Query Match 60.0%; Score 15; DB 149; Length 776;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcagtgatc 16
|||||
Db 618 CGCAGCAGCGCATC 604

RESULT 34
BG299490/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BG299490 792 bp mRNA EST 21-FEB-2001
HVSMEa0020F23f Hordeum vulgare seedling shoot EST library
HVCbNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0020F23f.
mRNA sequence.
BG299490
BG299490.1 GI:13086839
EST.
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 792)
Wing, R., Close, T. J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and
Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu

FEATURES
source
1..792
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEa0020F23f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCbNA0001 (Cold stress)"
/tissue_type="Seedling shoot"
/lab_host="TJc121"
/note="Vector: lambdaZAP; Site_1: EcORI; Site_2: XhoI"
BASE COUNT 123 a 229 c 235 g 183 t 22 others

Query Match 60.0%; Score 15; DB 175; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtcatcgccagcat 24
|||||
Db 459 GTGCATCGCCAGCAT 445

RESULT 35
BG309144
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BG309144 824 bp mRNA EST 22-FEB-2001
HVSMEC0004A20f Hordeum vulgare seedling shoot EST library
HVCbNA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEC0004A20f, mRNA sequence.
BG309144
BG309144.1 GI:13109991
EST.
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 824)
Wing, R., Close, T. J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and
Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu

Seq primer: AATTAACTCTACTAAAGG
High quality sequence stop: 657.
Location/Qualifiers

FEATURES

1. 824
/organism="Hordex vulgare"
/cultivar="Hordex"
/db_xref="taxon:4513"
/clone="HVSMEC0004A20f"
/clone_1lb="Hordex vulgare seedling shoot EST library
HVSMEC0003 (Etiolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP, Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 247 a 177 c 253 g 147 t
ORIGIN

Query Match 60.0%; Score 15; DB 152; Length 824;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcat 15
|||||
DB 801 GCCAGCAGGTGCAT 815

RESULT 36
AJ282984 843 bp mRNA EST 30-JUN-2000

LOCUS 4A3A-P2G5-R Anopheles gambiae immune competent 4A3A Anopheles
gambiae CDNA clone 4A3A-P2G5, mRNA sequence.

ACCESSION AJ282984
VERSION AJ282984.1 GI:6930863

KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae

REFERENCE 1 (bases 1 to 843)
Dimopoulos, G., Casavant, T. L., Chang, S., Scheetz, T., Roberts, C.,
Donohue, M., Schultz, J., Benes, V., Bork, P., Ansoorge, W., Soares, M. B.
and Kafatos, F. C.
Anopheles gambiae pilot gene discovery project: identification of
mosquito innate immunity genes from expressed sequence tags
generated from immune-competent cell lines
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)

TITLE Contact: Dimopoulos G
Fotis C. Kafatos Laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
Location/Qualifiers

COMMENT 1. 843
/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-P2G5"
/clone_1lb="Anopheles gambiae immune competent 4A3A"
/cell_line="Immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoRI; Site_2: NotI; sequenced from
forward priming site which reads from the 3' end of the
CDNA. The 4A3A is a directionally cloned and normalized
CDNA library that was constructed from the 4A3A cell line
oligo-T primed cDNA according to: Bonaldi, Lennon & Soares
(1996) : Normalization and Subtraction: Two approaches to
facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT 224 a 227 c 231 g 160 t 1 others

ORIGIN

BASE COUNT 224 a 227 c 231 g 160 t 1 others

Query Match 60.0%; Score 15; DB 104; Length 843;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcagcagcat 25
|||||
DB 512 TGCATGCCAGCAT 526

RESULT 37
BE731371 886 bp mRNA EST 15-SEP-2000

LOCUS 601565029f1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3840078 5',
mRNA sequence.

ACCESSION BE731371 GI:10145363
VERSION BE731371.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 886)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LLCM529 row: 1 column: 07
High quality sequence start: 3
High quality sequence stop: 776.
Location/Qualifiers

FEATURES

1. 886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3840078"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(c). Size-selected
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 151 a 279 c 271 g 185 t
ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcat 15
|||||
DB 785 GCCAGCAGGTGCAT 799

RESULT 38
BF684687 899 bp mRNA EST 22-DEC-2000

LOCUS 602141056f1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4302204 5',
mRNA sequence.

ACCESSION BF684687
VERSION BF684687.1 GI:11970095
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 899)
TITLE NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapsof@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM163 row: P column: 13
High quality sequence stop: 584.
Location/Qualifiers
1. 899
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4302204"
/clone_lib="NIH-MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site-1: XhoI; Site-2: EcoRI. CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 171 a 249 c 341 g 138 t
ORIGIN

Query Match 60.0%; Score 15; DB 168; Length 899;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgcagcagtgcat 15
|||||
Db 830 GCCGACGAGGTGCAT 844

RESULT 39
BE709927 116 bp mRNA EST 12-SEP-2000
LOCUS BE709927
DEFINITION IL3-HT0619-280600-192-H06 HT0619 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE709927
VERSION BE709927.1 GI:10098296
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 116)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-IL3-HT0619-280600-192-H06&t3=2000-06-28&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 57
High quality sequence stop: 115.
Location/Qualifiers
1. 116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0619"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site-1: SmaI; Site-2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and CDNA amplification were performed under low stringency conditions."

BASE COUNT 27 a 34 c 32 g 23 t
ORIGIN

Query Match 56.0%; Score 14; DB 139; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgcagcagtgca 14
|||||
Db 51 GCCGACGAGGTGCA 64

RESULT 40
BF876173 119 bp mRNA EST 17-JAN-2001
LOCUS BF876173
DEFINITION CM4-ET0097-111100-426-H09 ET0097 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF876173
VERSION BF876173.1 GI:12266303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 119)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM4&t2=CM4-ET0097-111100-426-H09&t3=2000-11-11&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 118.
Location/Qualifiers

FEATURES

source

1. 119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0097"
/dev_stage="Adult"
/note="Organ: Lung; tumor; Vector: puc18; Site: 1; SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from OPRESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 29 a 24 c 35 g 31 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 14; DB 170; Length 119;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 cagcaggtgcatcg 17
|||||
Db 7 CAGCAGGTGCATCG 20

RESULT 41

A259503/c

LOCUS

1M0407F22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0407F22 F, DNA sequence.

ACCESSION

A259503

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 178
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0407F22"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 33 a 46 c 47 g 52 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 14; DB 246; Length 178;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcatc 16
|||||
Db 115 GCAGCAGGTGCATC 102

RESULT 42

BF017303/c

LOCUS

BF017303 x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:3467679 3' similar to TR:070278 070278 MULTIPLE ENDOCRINE
NEOPLASIA TYPE 1 CANDIDATE PROTEIN NUMBER 18. ;, mRNA sequence.

ACCESSION

BF017303

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 196
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3467679"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 56.0%; Score 14; DB 143; Length 196;
 Best Local Similarity 100.0%; Pred. No. 6.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 agcaggtcagtcgc 18
 |||||
 Db 186 AGCAGGTGCATCGC 173

RESULT 43

LOCUS A1714341 206 bp mRNA EST 08-JUN-1999
 DEFINITION UI-R-AF1-aay-e-12-0-UI.s1 UI-R-AF1 Rattus norvegicus cDNA clone
 A1714341
 UI-R-AF1-aay-e-12-0-UI 3', mRNA sequence.
 A1714341
 EST
 A1714341.1 GI:5018141
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 206)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.iowa.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 M.B. Soares Lab Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLY-A=Yes.

FEATURES
 Source Location/Qualifiers

1..206
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AF1-aay-e-12-0-UI"
 /clone_lib="UI-R-AF1"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF1
 library is a normalized library constructed from 15 dpc
 rat atrioventricular (AV) canal. The tag is a string of 5
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa.
 TAG_LIB=UI-R-AF1
 TAG_TISSUE=AV canal at 15 dpc
 TAG_SEQ=CAAGG"
 BASE COUNT 51 a 58 c 41 g 56 t
 ORIGIN

Query Match 56.0%; Score 14; DB 246; Length 206;
 Best Local Similarity 100.0%; Pred. No. 6.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 ggcagaggtcagtc 16
 |||||
 Db 149 GCAGCAGGTGCATC 136

RESULT 44

LOCUS BB563629 219 bp mRNA EST 29-NOV-2000
 DEFINITION BB563629 RIKEN full-length enriched, 18 days embryo Mus musculus
 cDNA clone 1110002M02 5', mRNA sequence.
 BB563629
 BB563629.1 GI:11454521
 EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 219)
 AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.,
 Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T., Hodyama,Y.,
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Kono
 ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
 Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
 Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A.,
 Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
 ,T., Toyota,T., Watanabe,K., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshiki,A., Yamatsu,M., and Hayashizaki,Y.
 RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
 UNpublished (2000)
 TITLE Contact: Yoshihide Hayashizaki
 JOURNAL Genome Exploration Research Group, Life Science Tsukuba Center,
 COMMENT Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 URL:http://genome.rtc.riken.go.jp/
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasaki
 ,N., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
 Thermostabilization and thermoactivation of the thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Katsunari,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 ,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)

FEATURES

Source

Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

Location/Qualifiers

1..219
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="1110002M02"
 /clone_lib="RIKEN full-length enriched, 18 days embryo"
 /dev_stage="18 days embryo"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: SstI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'

Search completed: October 9, 2001, 15:15:31
Job time: 13653 sec

GAGAGAGAGCGCCGCAACTGAGTTTTTTTTTTTTTTVN 3'}, cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'

BASE COUNT
ORIGIN

46 a 76 c 46 g 51 t

Query Match

Best Local Similarity 56.0%; Score 14; DB 161; Length 219;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgca 14
|||||

DB 57 GCGCAGCAGGTGCA 70

RESULT 45

AM326327 226 bp mRNA EST 09-JUL-2000
LOCUS 18754 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

AM326327
AM326327.1 GI:6762248

EST
KEYWORDS

SOURCE

ORGANISM

Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
1 (bases 1 to 226)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.,
and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine

Unpublished (2000)
Contact: Smith TPL

JOURNAL

COMMENT

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 20
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCAGCAGC
Plate: 10 Row: K Column: 8

Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers

1..226
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/issue_type="pooled"
/lab_host="DH10B"

/note="Vector: PCMW SPORT6; Site 1: XbaI; Site 2: XhoI;
library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

FEATURES
source

BASE COUNT

40 a 71 c 75 g 40 t

ORIGIN

Query Match

Best Local Similarity 56.0%; Score 14; DB 114; Length 226;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgca 14
|||||

DB 173 GCGCAGCAGGTGCA 186

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:45:48 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25
Sequence: 1 gcaggtgcatgcgcacattcgat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 207944

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:**

1: gb_da1:**

2: gb_da2:**

3: gb_da3:**

4: gb_in1:**

5: gb_in2:**

6: gb_in3:**

7: gb_om:**

8: gb_ov:**

9: gb_pat1:**

10: gb_pat2:**

11: gb_ph:**

12: gb_pl1:**

13: gb_pl2:**

14: gb_pl3:**

15: gb_pl4:**

16: em_ba1:**

17: em_ba2:**

18: em_fun:**

19: em_htgo_hum:**

20: em_htgo_inv:**

21: em_htgo_rod:**

22: em_htg_hum1:**

23: em_htg_hum2:**

24: em_htg_hum3:**

25: em_htg_hum4:**

26: em_htg_hum5:**

27: em_htg_hum6:**

28: em_htg_hum7:**

29: em_htg_hum8:**

30: em_htg_inv1:**

31: em_htg_inv2:**

32: em_htg_other:**

33: em_htg_rod:**

34: em_hum1:**

35: em_hum2:**

36: em_hum3:**

37: em_hum4:**

38: em_hum5:**

39: em_hum6:**

40: em_hum7:**

41: em_hum8:**

42: em_om:**

43: em_or:**

44: em_ov:**
45: em_pat:**
46: em_ph:**
47: em_pl:**
48: em_ro:**
49: em_sts:**
50: em_sy:**
51: em_un:**
52: em_v1:**
53: gp_sts1:**
54: gp_sts2:**
55: gp_sts3:**
56: gp_sy:**
57: gp_un:**
58: gp_v1:**
59: gp_v12:**
60: gp_v13:**
61: gp_v14:**
62: gp_v15:**
63: gp_v16:**
64: gp_v17:**
65: gp_v18:**
66: gp_v19:**
67: gp_v20:**
68: gp_v21:**
69: gp_v22:**
70: gp_v23:**
71: gp_v24:**
72: gp_v25:**
73: gp_v26:**
74: gp_v27:**
75: gp_v28:**
76: gp_v29:**
77: gp_v30:**
78: gp_v31:**
79: gp_v32:**
80: gp_v33:**
81: gp_v34:**
82: gp_v35:**
83: gp_v36:**
84: gp_v37:**
85: gp_v38:**
86: gp_v39:**
87: gp_v40:**
88: gp_v41:**
89: gp_v42:**
90: gp_v43:**
91: gp_v44:**
92: gp_v45:**
93: gp_v46:**
94: gp_v47:**
95: gp_v48:**
96: gp_v49:**
97: gp_v50:**
98: em_da3:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	9	AR029499
2	25	100.0	1041	9	AR034916
3	25	100.0	1084	9	A11530
4	25	100.0	1121	10	E00893
5	25	100.0	5793	2	ECOBIO
6	25	100.0	5872	9	A38246
7	25	100.0	5872	9	A38251
8	25	100.0	5872	9	A93674

```

9      100.0  5872  9  A93679
10     100.0  5872  9  AR101809
11     100.0  5872  9  AR101810
12     100.0  11022 1  AE000180
13     100.0  13501 1  AE005258
14     100.0  297816 2  AP002553
15     19  76.0  3603 3  PSEHRIA
16     18  72.0  2923 3  MAV250020
17     17  72.0  5526 2  AF250776
18     17  68.0  479 54  G01446
19     17  68.0  53067 65  AC019940
20     17  68.0  153870 63  AC013431
21     17  68.0  183937 71  AC037435
22     17  68.0  327446 4  AE003500
23     16  64.0  1186 2  AF146614
24     16  64.0  4674 89  AK024397
25     16  64.0  11910 56  AF206717
26     16  64.0  49587 6  DMBH48C10
27     16  64.0  60573 65  AC018267
28     16  64.0  82289 13  ATAC009895
29     16  64.0  117968 92  HS272E8
30     16  64.0  140825 13  AP002817
31     16  64.0  146081 13  AP001366
32     16  64.0  154306 71  AC034114
33     16  64.0  177380 85  AC002457
34     16  64.0  180700 76  AC084136
35     16  64.0  183439 4  AC007809
36     16  64.0  232744 5  AE003705
37     16  64.0  237119 65  AC017740
38     16  64.0  289090 4  AE003424
39     15  60.0  1389 14  NTEPCL1
40     15  60.0  1713 14  NTEPCL1
41     15  60.0  1907 93  HSSYNTR05
42     15  60.0  2616 85  AB048915
43     15  60.0  3369 13  AF274670
44     15  60.0  3369 13  AF274671
45     15  60.0  3677 3  PSEIAAHM

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    source          1..1041
                     /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatgcgcacgacattcgat 25
|||||
DB 84 GCAGGTGCATCGCCAGCATTTTCAT 108

```

RESULT 2

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
    source          1..1041
                     /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatgcgcacgacattcgat 25
|||||
DB 84 GCAGGTGCATCGCCAGCATTTTCAT 108

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Escherichia.
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
    source          1..1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
                     /gene="B10B"
                     /gene="B10B"
                     /gene="B10B"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490218"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWLTISOYTRLEPKRPLDILFEAQOVHROHFDPPQOVOST
                     LSIKTAGACPEOCKYCROSKRTGTLAEARLKEVEQVESARRAKAAGSTRPCGAAM
                     KNHERDMPTLEDMOVGVAKAMGLEACMTGTLSESOAQLANAGLDYVNHNDTSPEF
                     YGNTITTRTYOERLDLETKVRDAGIKVCSGGIVIGETVKDRAGILLQIANLPTPES
                     VPIINMLVKVGTPLADNDVDADFETRTIAVARIMMPTSYVRLSAGREDMMNQTQAMC
                     FMAGANSIFYGCKLLTTPNPERDKDLQLEFRKGLNPDQTAVIAGNEDQRIEQALMT
                     PTDDEVYNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatgcgcacgacattcgat 25
|||||

```



```

RBS
105. .119
/standard_name="BIOB_RBS NO.9"
/evidence=experimental
117. .1157
/gene="BIOB"
117. .1157
/gene="BIOB"
/codon_start=1
/translation_label=11
/number=1
/evidence=experimental
/product="BIOB_TIN SYNTHASE"
/protein_id="CAO2325.1"
/db_xref="GI:2294845"
/translation="MAHRRPMTLSQVTELEFEKPLDLLFEAQOVRHOFHPROVOST
LLESLKQACEDCKCYKPOSSRYKTELEARIEMVEOYLRSAKAKAAGSTRCCMAA
KNHEDRDELDKMOGVOSYKAMDLACMHLGTLSSQOYLRANAGLDYVNHNLDTSPBE
YGNITITTRYOERLDLTKVRDAGIKVCSGIVGLCTVYDRAGLLDLQANLPTPEE
VPIINMLVKKGTPPLADNDVDYAFDFTIRIAVARIMPTSVRLSAGREOMQOTAMQ
FMAGASIFKVGKLLTTNPEDKDLQLFRTKLGINPQQTAVLAGNDQQORLEQALM
PDTDEVYNAAL"
2284. .2297
/standard_name="BIOC_RBS"
2295. .3050
/gene="BIOC"

```

```

/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"

```

```

RBS
/standard_name="BIOA_RBS"
3742. .3752

gene
/standard_name="BIOA_RBS"
3750. .5039
/gene="BIOA"
3750. .5039

CDS
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5

```

[illegible]

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDDDKLPHRHVFNCGMGVDDGNISPHLAMDDVPAGTK
SFVAVCYDDAPDTGSGMMHWVNNLPADTRVLPQFGSLVAMPDGVLTQTRDCKTG
YDGAAPPKGETHRIFYTHALDIERIDVDEGASGAMVGHNVHFHSLASSTAMFS"
5583..5605
5583..5644
stem_loop
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gccagtgatcgccagcatcgcattcgat 25
|||||
Db 200 GCAGTGCATCGCCAGCATTTGCAT 224
RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
COMMENT
other publication PL 308301 950724
other publication CA 2145400 940414
other publication AU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9500809 950913
other publication FI 951547 950331
other publication JP 8501694T 960227.
FEATURES
source Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKINAADARADALRRYPVAGAGRWLVADROYLNF
SSNDYLGSHFPIILRAMQGAKEFGTSGSGHVSGLSYHQALEELAEMLGYSRA
LFIISGFANAVIAAMAKEDRIADRLISASLSEASLSPSQLRRRAHNDVTHLAR
LIASPCGQMVVTEGVFSMDGDSAPLAEIDQVTOOHNGWMLVDDAHGTGVIGEGRG

SCWLOKYPPELLVTFGKFGVSGAAVLCSTVADYLLQFARHLIYSTSMPAQAOL
RASLAVTRSDGGRAREKLAALITRPAGVQDLPFTLADSCSAIOPLIVGDNRAIOL
AEKLKQGCWTAIRPTVPAGTARLRLLTAAHMQDDIDRLLEVLHNGN"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGDTVEGKTVASCALLQAAKAGRTAGYKPPVASS
EKTEPGLNSDALALQRRNSSLQDYATVNPYFAEPSPHLISAQGRPIESLVMASG
LRALRQADWVLEAGCGMTPLSDTFEFADWVQEOPLVILVGVKLCINHAMLTA
OYIOHAGLTLAGWVANDVTPPGKRRAEYMTTLTRMIPAPLIGIPLMIAENPENATGK
YINLAVDASTLIGTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gccagtgatcgccagcatcgcattcgat 25
|||||
Db 200 GCAGTGCATCGCCAGCATTTGCAT 224
RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
FEATURES
source Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1

```

/transl_table=11
/evidence=experimental
/product="Biotin SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRNLTLSOVELPEFKPLDILFEACQVHROHPFPROVOST
LSTKTCGCEPDCYCCOSSRYKGLAEAELEMEYOLESAKAKAGSTRFCMGAW
KNRHRMPYLEOVQCVKAMGLEACHTGLTSQAKRLANAGLDYHNHNDLPTSEF
YGNITTRTTOERLDLEKVDAGIKVCSGIVGLGKVDAGLILLOLANLPTPES
VPIMLVKVGTPLADNDVDADPELFTIIVARIMPTSYRLSAGREONBOTOAMC
FMAGANSIFGCKLTPNPEEDKDLDFKRLGLNPQATVLAGDNQQRRLQALMT
PDTDEYYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQATIAAAGRAAAHYEQHADLQROSADALLAMIPORRYT
HYLDACGCGGMSRHRERHRAQVYATLDLSPMIVOROKDADHYLAGDIESPLATA
TFLDASNLAVOMCGNLSTALRELYVRPKCVATTTVCGSLPELHOAWADBRP
HANRFLPEDEIROSINGVYCHHIOPIITLFPDALSAMSLGIGATHLHESRPRIL
TRSOLORIOLAMPQOQGRFLIYHLGLVLAHE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/ec_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADQRHIMHPYTSMTSPVYEVVSAECELLISDGR
LVDGMSWMAIIGHYNHOLNAKMSQIDAMSHVMFGGITAPATELCRLIVAMPPOP
LECVLADSGSYAVEAMKALOYQWAGEARORFLTPRNGYHGDTPGAMVCVQDPS
MHSILKGYLPELFAPOSIMDGEDMDERDVGARLMAHRHEIAVITETPYQAG
GMRVHPEMIKRIKICDRREGILADIATGFGTGKLFACBAELIAPDILGKAL
TGGWTLNATLITRIVATISNGEAGCFHGPTEGNCPLACAAANASIALLESQMOO
QVADIEVQREOLARDAEMVADRVLGATGVETTHPVNMAALQKFEFVQGWIRP
FEKLIYMPPIITLPEQDLRLIAVNAVODETFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLKLRHVPNGMGYDGDNISPHLAMDVYPAIGK
SFVVTGYDPAPGTSGSMHWVYVLPADTRVLPQGGSLVAMPDGYLQNRTPGKIG
YDGAAPKGETHRIETVHALDIRIVDGSAGSAMGVNHHISLASASITAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcacgcgcacattcgat 25
Db 200 GCAGGTGCATGCCACATTTCGAT 224

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
BIRCH, O. and BRASS, J.
BIOLOGICAL METHOD OF PRODUCING BIOTIN
PATENT: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/ec_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOEKINALDARRAADLRPRYPVAGAGRWIVADROYLNF
LFIISGFPAQNAVIAAMAKEDRIADRSHASLSFASLSFQLRFRFANVTHLAR
LLASPCGQDQVTEGVFSMDGSDAPLARIQVYTOQNGMVLWDAGHGVIGEDGRG
SCMLQKVRPELVITFGKGVSGAVLIGSTVADYILOFARHLIYSTSPPAQOAL
RASLAVIRSDGEGDAREKLAALITFRRAQVODLPTLADSCAIOPLIYDMSRALQL
AEKLRQCGCWVTAIRPPTVAGTARLRILTAAHEMODIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/ec_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTR SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFYVGTPEVKGTVASCALLQAKKAGYPRAGTRPPVAGS
EKTPEGRNSDALAQRNSSQIDLYATVNPYTFABETSPHITSQEGRPISLVSAG
LRAIRQADWVTVBGAGWFTPLSDTFPAWVQOQLPVILVGVKLGCIHMLT7A
QVIOHAGLTLAGVANDVTPPGKRHAETMTLTTRMIPAPLIGELIPWLAENBNAATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t


```

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
VINAFAVDASTLIGFTSRL"

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatccgcagcatcgcattgc 25
    ||||||||||||||||||||||||||||
Db 200 GCAGGTGCATCCGCAGCATTTTCGAT 224

RESULT 10
LOCUS      ARI01809      5872 bp      DNA
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION  ARI01809
VERSION    ARI01809.1 GI:12812607
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES   Location/Qualifiers
            source 1..5872
            /organism="Unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
VINAFAVDASTLIGFTSRL"

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatccgcagcatcgcattgc 25
    ||||||||||||||||||||||||||||
Db 200 GCAGGTGCATCCGCAGCATTTTCGAT 224

RESULT 11
LOCUS      ARI01810      5872 bp      DNA
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION  ARI01810
VERSION    ARI01810.1 GI:12812608
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES   Location/Qualifiers
            source 1..5872
            /organism="Unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
VINAFAVDASTLIGFTSRL"

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatccgcagcatcgcattgc 25
    ||||||||||||||||||||||||||||
Db 200 GCAGGTGCATCCGCAGCATTTTCGAT 224

```

```

RESULT 12
LOCUS      AE000180
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION  AE000180 U00096
VERSION    AE000180.1 GI:1786988
KEYWORDS
SOURCE     Escherichia coli K12.
ORGANISM   Escherichia coli K12.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.

REFERENCE  1 (bases 1 to 11022)
AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
            9742617
            MEDLINE
            PUBMED
            9278503
            2 (bases 1 to 11022)
            Blattner,F.R.
            Direct Submission
            Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 11022)
            Blattner,F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 11022)
            Plunkett,G. III.
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
            30332 [e-mail: mark@ambr.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated: this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.

FEATURES
source      Location/Qualifiers
            1..11022
            /organism="Escherichia coli K12"
            /strain="K12"
            /sub_strain="MG1655"

```

```

repeat_region
/db_xref="taxon:83333"
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhc"
/note="b0772"
complement(147..1430)
/gene="ybhc"
/function="orf: Not classified"
/note="f427: 98 pct identical to fragment YBHC_ECOLI
SM:P46130 (300 aa) but contains 127 additional C-terminal
residues"
CDS
/transl_table=1
/codon_start=1
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVSRLLALAFGVTLTACSSPPDQPSDQAPGSSRPI
LSAKAQNPDQOHFASLTPGAAMNPSPITIPADPPVYGVAGTGVHTT1QAADV
AALIKRNKROYIAVMGEYGVYVPAAGGITLGTGKPLDVKIGSLDGMSPA
DMRHVDNPRGKWMGKPRMYMYSKSDSGVCSAVESQNNGLDQLNLTENT
NSYIEGVDIVSGAVEDNTEFRVYNSRTQEAFAATLSNIYGEFLAVNSRF
AFEDGVAQLGRSLDVANDANTNGVYIRDSAINEGFNTAKPWADAVISNRPAGTGSVD
DNDEIQNLNDINYNRMWEXNNRGVSKVVAEAK"
complement(1507..1535)
/note="factor Sigma70: predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhb"
/note="b0773"
complement(1582..2058)
CDS
complement(1582..2058)
/gene="ybhb"
/function="orf: Unknown"
/note="f158: 99 pct identical to YHB_ECOLI SM: P12994"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDCKLPHRIHFGMGVGDGNTSPLAMDDVAGTK
SEFVTCYDPDAPATGSGMHHVYVNLPAIDRVLPQSGSLVAMPDGVLTQTRDEKGTG
YGAAPKGEKTHRYETFAHLDIERIDDEGSGAMVGNVHFHSLASASTYAMS"
complement(2108..2124)
/note="central position to predicted promoter:85.5"
/bound_moiety="Rhas predicted site"
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/EC_number="2.6.1.62"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="f429: 100 pct identical to BIOA_ECOLI SM: P12995"
/codon_start=1
/transl_table=1
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDIAFDORNIWHPTSTPLPYVVSAGGCELLISDGR
LVDSQSMMAIHGYNHNPOLNAAMKSOIDANSHVFGGITHAPRIETCRIVAMTPPO
LIECVLADSGVAEVAEMKMALOYWAKGAEORFLPFRCGHDYTGANSVCDPDS
MHSILMGYULPENLFAAPAROSMGDEMRWGRARLMAHRRHIAVITPTVOGAG
GMRYHPHEMLKRIKICDREGILLIDETIATGERTGLFACHEAETAPDILDGLAL
TGCTMTSATITTRVAVETISNAGCFMNGPFGMNGPACAAANSIALIEGSDMOO
QVADIEVLKEQLAPADAMVADVRLGIVGETHPVMAALCKFVEGQVWLRP
FGKILMPYIILPOLQRLTAAVNVAODETFEQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70: predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

protein_bind
/bound_moiety="bioB documented site"
3411..3450
/note="central position to predicted promoter: -20;genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
3413..3441
/bound_moiety="BioB documented site"
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o346: 99 pct identical to BIOB_ECOLI SM: P12996"
/codon_start=1
/transl_table=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPMTLSQVTELEFEKPLDLLFPAQGVHROHPDPROVOYST
LILRTGACPEDECKYCPQSSRYKTGLAEERLMEYVLESARAKAGSRPFMGAMW
KNPHERMPYLEQWGVKAMGLACACTLIGTSQOARLANGLYNNIDTSEB
YGNITTRTYOERLDTEKVRDAGIKYCSGICYGCTVYDRAGLLQLANLPTPES
VPIMLVKVGCTPLADDNDAPRPFITIVARIIMPTSTVRLSAGREMNQEQIAC
FMAGANSIFVGCILITTPNEBDKDLQFLKGLINPOVTVAGLDNEQDRLBOALMT
PDDEYTNAAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o384: 100 pct identical to BIOF_ECOLI SM: P12998"
/codon_start=1
/transl_table=1
/product="8-aminoc-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWOKIRNAALDARRAADALRRKYPVAGAGRWLVADDROYLNF
SSNDYGLSHHPQILIRAWOGAEOFGISGSGHVSQVYVHQALEELAEELGYSRA
LFIISFGAANOAVIAIAMAKEBRIADRLSHASLEASLSPOLRRFAHNDYTLAR
LIASPCGOOMVATEGVFMSDGBSAPLAEIQOVTQOCHNGLMWADAHGVYIGEGRG
SCWLAKVKEPELLVTRGKFGVSGAAVLCSSIVADYLLQFAHLLYTSMPPAQAL

gene
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o384: 100 pct identical to BIOF_ECOLI SM: P12998"
/codon_start=1
/transl_table=1
/product="8-aminoc-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWOKIRNAALDARRAADALRRKYPVAGAGRWLVADDROYLNF
SSNDYGLSHHPQILIRAWOGAEOFGISGSGHVSQVYVHQALEELAEELGYSRA
LFIISFGAANOAVIAIAMAKEBRIADRLSHASLEASLSPOLRRFAHNDYTLAR
LIASPCGOOMVATEGVFMSDGBSAPLAEIQOVTQOCHNGLMWADAHGVYIGEGRG
SCWLAKVKEPELLVTRGKFGVSGAAVLCSSIVADYLLQFAHLLYTSMPPAQAL

CDS
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/EC_number="2.6.1.62"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="f429: 100 pct identical to BIOA_ECOLI SM: P12995"
/codon_start=1
/transl_table=1
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDIAFDORNIWHPTSTPLPYVVSAGGCELLISDGR
LVDSQSMMAIHGYNHNPOLNAAMKSOIDANSHVFGGITHAPRIETCRIVAMTPPO
LIECVLADSGVAEVAEMKMALOYWAKGAEORFLPFRCGHDYTGANSVCDPDS
MHSILMGYULPENLFAAPAROSMGDEMRWGRARLMAHRRHIAVITPTVOGAG
GMRYHPHEMLKRIKICDREGILLIDETIATGERTGLFACHEAETAPDILDGLAL
TGCTMTSATITTRVAVETISNAGCFMNGPFGMNGPACAAANSIALIEGSDMOO
QVADIEVLKEQLAPADAMVADVRLGIVGETHPVMAALCKFVEGQVWLRP
FGKILMPYIILPOLQRLTAAVNVAODETFEQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70: predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

promoter
/note="factor Sigma70: predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

protein_bind
/bound_moiety="bioB documented site"
3411..3450
/note="central position to predicted promoter: -20;genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
3413..3441
/bound_moiety="BioB documented site"
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o346: 99 pct identical to BIOB_ECOLI SM: P12996"
/codon_start=1
/transl_table=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPMTLSQVTELEFEKPLDLLFPAQGVHROHPDPROVOYST
LILRTGACPEDECKYCPQSSRYKTGLAEERLMEYVLESARAKAGSRPFMGAMW
KNPHERMPYLEQWGVKAMGLACACTLIGTSQOARLANGLYNNIDTSEB
YGNITTRTYOERLDTEKVRDAGIKYCSGICYGCTVYDRAGLLQLANLPTPES
VPIMLVKVGCTPLADDNDAPRPFITIVARIIMPTSTVRLSAGREMNQEQIAC
FMAGANSIFVGCILITTPNEBDKDLQFLKGLINPOVTVAGLDNEQDRLBOALMT
PDDEYTNAAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o384: 100 pct identical to BIOF_ECOLI SM: P12998"
/codon_start=1
/transl_table=1
/product="8-aminoc-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWOKIRNAALDARRAADALRRKYPVAGAGRWLVADDROYLNF
SSNDYGLSHHPQILIRAWOGAEOFGISGSGHVSQVYVHQALEELAEELGYSRA
LFIISFGAANOAVIAIAMAKEBRIADRLSHASLEASLSPOLRRFAHNDYTLAR
LIASPCGOOMVATEGVFMSDGBSAPLAEIQOVTQOCHNGLMWADAHGVYIGEGRG
SCWLAKVKEPELLVTRGKFGVSGAAVLCSSIVADYLLQFAHLLYTSMPPAQAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagtgacgcgcacgcatcgat 25
|||||
Db 3576 GCAGGTGCATCGGCACATTTGCAT 3600
|||||

RESULT 13
AE005258
LOCUS AE005258 13501 bp DNA
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155
ACCESSION AE005258 AE005174

```

```

VERSION      AE005258.1 GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        Nature 409 (6819), 529-533 (2001)
JOURNAL      21074935
MEDLINE      11206551
PUBMED
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
              Location/Qualifiers
              1. 13501
                 /organism="Escherichia coli O157:H7 EDL933"
                 /strain="EDL933"
                 /serotype="O157:H7"
                 /db_xref="taxon:155864"
                 /note="enterohaemorrhagic"
                 <1..7576
                 /note="O-island #36. Region of the EDL933 chromosome not
                 homologous to E. coli K-12 MG1655; cryptic prophage
                 CP-933K; includes one copy of the 13 bp direct repeat that
                 flanks the prophage"
                 66..665
                 /gene="10mk"
                 /note="20981"
                 66..665
                 /gene="10mk"
                 /function="putative membrane; Other or unknown (Phage or
                 prophage Related)"
                 /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                 residues 1 to 199 of 199 from GenPept 118 :
                 g117532789|gb|AA63231.1|AF151091_2 (AF151091) Lom
                 |prophage P-E1bA|"
                 /codon_start=1
                 /transl_table=1
                 /product="putative outer membrane protein of prophage
                 CP-933K"
                 /protein_id="AAG55137.1"
                 /db_xref="GI:12513752"
                 /translation="MRKYCAILSAICLAVSGVPAMASEHOSITAGYLHASTDAPG
                 SDDLNGIWKYRYEPTDLGLITSPSTANADECKTISTDRHEDYVRRRWISVAG
                 PSVRNEMFSAVAMAGVAVSRVSTSGDYEVTDNKKRTDVLGTGSDARTSVTSIAW
                 GAGVOFNPTESSAVADVAVEXSGSDWRTDFIVGYGK"
                 719..2041
                 /gene="20982"
                 719..2041
                 /gene="20982"
                 /function="putative structure; Structural component (Phage
                 or Prophage Related)"
                 /note="Residues 164 to 440 of 440 are 68.79 pct identical
                 to residues 381 to 645 of 645 from GenPept 118 :
                 g114585437|gb|AAD25464.1|AF125520.59 (AF125520) putative
                 tail fiber protein [bacteriophage 933W]"
                 /codon_start=1
                 /transl_table=1
                 /product="putative tail component of prophage CP-933K"

gene
CDS
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVIOGLVTKDGACKPIQNTCIOLKARNSTXYVNVTVASE
NDEGRISMDVEYQCYVTVLLVEGFPSPSHACTISYVEDSOPGLINDFLGAMEDAR
PALRFRFEQWVEEARHAEKRNKAGETARNAKISASAKASANAPTSADPASE
SAROAASAASAKSESSSSASAAKASESIQSATDALSKKTESAAGNARARA
TSTTEKARSASQASAEOSRIADDAVNRIPYVGPGRGEPGPGPGPGDKE
KDTGTGAGATGTRGPGPGPGPGDKEGTEGIGTGNAGPGPGPGPGDGA
GPGPGKGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG
TGTFTLDPKDATQVGMQHLQVRRGDDGPGMDVYKGLDVSQDTRTGE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from GenPept 118 :
g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [bacteriophage 933W]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MRKAVIOALENDNNILKRLQSLGCGKHDCCENGSLTAQLR
LGPADLLESDENCIIPEDRVITQVYLLDADKKIQCVVRPLQILRADGTWENIGMK"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 :
g116960367|gb|AA633527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MSPIRTFTHNSVNIYVSSPCQTVSFAGEYELAKIDKPTPLF
LQEPNPERYKDEVPYVNTKQHPYLIDNVNNAIRISDRKIGIFVGDSSVQKRAF
KLEDFENVMYIREDVDVFSYDRLSDIYHDIICQRLTEDEKRDREYLLNLEELR
EISRAQDSISIMYAKKRNHAPDFEERLALFKAGEIIFCYNTKNGHISFGGCIYLD
MDMLTGLGTLITVAPDQISNHYDRNDNSVIENTSALITVNSNHPALLEGISFMH
AHFYDGLGKGVKRYFNFPTLHNINHCDFIEFNHRIITNTSQTCSM"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPISLDSNPNFSPAGYSAIPAPNRAENAYADYLDIGRIPL
SAADLSNVEYSVIRAVHDSRSRLIDQHTVDAIGTIVDALSRSGTDFDASVGIINEX
VHIGCTIKRNEYELNESSVKRIDIQSLTCKELLEYVGOEPIIPICEAENENPERX
VSFSAADPTDSYEMSQWEGLIHETIHHVYGSQSDSNIELEGPTEILARRAODELG
WSVPDFKGYAEPREARHLRLNLNALRQAARHENEKRAFERLGTISDYKVASPDT
KRSVANSNGYGFQIQHDPGGLAINDNLQDANQIOLYAGAPYIFFGVDVKNHOD"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAC55142.1"
/db_xref="GI:12513757"
/translation="MISPSYVNLGCSWNSITRLTSPDRNIVLSVDAVHSDNGAUV
KVGNRTRYVATONKPCVTRPRESHSGCTTLLHRLGMPKEISRIEVLMLNAPSVAM
FTTSQSEVYSEVRCFNOYAGASAEKTYGNGDIIIGIRMKINGESILNTSPDAO
AEHATYDFRLEOKGILLFVDTETVLDRAKNERNPIDISSVNSDRSMESSQIMQ
5462..7160
/gene="20990"
/length="7160"
/gene="20990"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAC55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNLVHOSRSSMSDPIESLVKASSVQMIKNNQRLP
QGDHNIYDIEALDKIGSTETGRVLLNAISIRLSEYVILHNSRLCVAHRD
IDAHNRGTSDPHCNLNAVEYPCGEGISVDFPATIVFHEHLHFNHNGERLKEVS
SRAESKYSPLLEEARVYGLAFSEVISEKPHHEIIGMPRTSYPSDALIHDDNT
VSLGFOOVRHLHPLT"
complement(7668..8144)
/gene="ydhb"
/note="20992"
complement(7668..8144)
/gene="ydhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
M61655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPHRHVNGMGYDGNISPHLANDVPAGTK
SEVVCYDPDAETGSGMMWVYVNLPAOTRVLPQGFSGLVAMPDGLVOTRTEGKTC
IDGAPPKGETHRVIFVHALVERIDVDEGASGAMGVNHFHSLASATIMFS"
complement(8203..9492)
/gene="b10a"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaagtcacgcacagcattcgat 25
|||||
Db 9662 GCAGTGCAATGCCACATTTCGAT 9666

```

RESULT 14

AP002553 297816 bp DNA BCT 07-MAR-2001

LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.

DEFINITION AP002553 BA000007

ACCESSION AP002553.1 GI:13360211

VERSION

KEYWORDS

SOURCE

Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)

DNA.

Escherichia coli O157:H7

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE

1 (sites)

ORGANISM

Escherichia coli O157:H7

gamma subdivision; Enterobacteriaceae;

REFERENCE

1 (sites)

AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 1113432
PUBMED 20575196
REFERENCE 2 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W. and Streif, W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES Location/Qualifiers
 1..5526
 /organism="uncultured bacterium pcosHE2"
 /db_xref="taxon:143797"
 /clone="pcosHE2"
 /note="unknown organism, cosmid clone derived from environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-bioA intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLDCDKLPHRHVFGMGYDGDNISPLAMDVPAGTKSFVVCYDDAPGTGSGMHVYVNLPAIDRVLPQCGSGLVAMPDGVLOTPTDCKAGYCGAAPKGETHRYIFTHALDIRIDVDEGASGAMVGFNVHSLASASITAMFS"
 complement(587..1876)
 /gene="bioA"
 /complement(587..1876)
 /gene="bioA"
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAEPDRIHMPYTSMTSPLFVYVVSAGCECELLSDGRKLVNGSSWMAIHCYNHNPOLNAMKSOIDAMSHVMFGITTHPALECKLIVAMPPOP LKCYFLADSGSAVEYAMKALDYOAKKEARQITTFNGYHGDPFGMSYCDSDNS MHSIWMGYLPENLFAPOSRMDGWDENDEMGFALMAHHEIAVYIEPIYOGAG GMRVYHEMLKRIKICDREGILLADELATGGRGKLFACHEAVIADIIICLKAL TGGTMTLSATLTREVAETISNGEACFMHGPTFMGNPLACAANAASIALLESQMOQ OVADIEVOLREOLAPARDAEMVADVILGALGVETTHVNNAAIÖKFEVEQGVIRP RGLTIYMPRYIILLPOOLRL/AAVNRVAVDDETFECQ"

gene
 CDS
 /gene="bioA"
 /complement(587..1876)
 /gene="bioA"
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAEPDRIHMPYTSMTSPLFVYVVSAGCECELLSDGRKLVNGSSWMAIHCYNHNPOLNAMKSOIDAMSHVMFGITTHPALECKLIVAMPPOP LKCYFLADSGSAVEYAMKALDYOAKKEARQITTFNGYHGDPFGMSYCDSDNS MHSIWMGYLPENLFAPOSRMDGWDENDEMGFALMAHHEIAVYIEPIYOGAG GMRVYHEMLKRIKICDREGILLADELATGGRGKLFACHEAVIADIIICLKAL TGGTMTLSATLTREVAETISNGEACFMHGPTFMGNPLACAANAASIALLESQMOQ OVADIEVOLREOLAPARDAEMVADVILGALGVETTHVNNAAIÖKFEVEQGVIRP RGLTIYMPRYIILLPOOLRL/AAVNRVAVDDETFECQ"

gene
 CDS
 /gene="bioB"
 /complement(587..1876)
 /gene="bioB"
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAEPDRIHMPYTSMTSPLFVYVVSAGCECELLSDGRKLVNGSSWMAIHCYNHNPOLNAMKSOIDAMSHVMFGITTHPALECKLIVAMPPOP LKCYFLADSGSAVEYAMKALDYOAKKEARQITTFNGYHGDPFGMSYCDSDNS MHSIWMGYLPENLFAPOSRMDGWDENDEMGFALMAHHEIAVYIEPIYOGAG GMRVYHEMLKRIKICDREGILLADELATGGRGKLFACHEAVIADIIICLKAL TGGTMTLSATLTREVAETISNGEACFMHGPTFMGNPLACAANAASIALLESQMOQ OVADIEVOLREOLAPARDAEMVADVILGALGVETTHVNNAAIÖKFEVEQGVIRP RGLTIYMPRYIILLPOOLRL/AAVNRVAVDDETFECQ"

gene
 CDS
 /gene="bioC"
 /complement(587..1876)
 /gene="bioC"
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAEPDRIHMPYTSMTSPLFVYVVSAGCECELLSDGRKLVNGSSWMAIHCYNHNPOLNAMKSOIDAMSHVMFGITTHPALECKLIVAMPPOP LKCYFLADSGSAVEYAMKALDYOAKKEARQITTFNGYHGDPFGMSYCDSDNS MHSIWMGYLPENLFAPOSRMDGWDENDEMGFALMAHHEIAVYIEPIYOGAG GMRVYHEMLKRIKICDREGILLADELATGGRGKLFACHEAVIADIIICLKAL TGGTMTLSATLTREVAETISNGEACFMHGPTFMGNPLACAANAASIALLESQMOQ OVADIEVOLREOLAPARDAEMVADVILGALGVETTHVNNAAIÖKFEVEQGVIRP RGLTIYMPRYIILLPOOLRL/AAVNRVAVDDETFECQ"

gene
 CDS
 /gene="bioD"
 /complement(587..1876)
 /gene="bioD"
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAEPDRIHMPYTSMTSPLFVYVVSAGCECELLSDGRKLVNGSSWMAIHCYNHNPOLNAMKSOIDAMSHVMFGITTHPALECKLIVAMPPOP LKCYFLADSGSAVEYAMKALDYOAKKEARQITTFNGYHGDPFGMSYCDSDNS MHSIWMGYLPENLFAPOSRMDGWDENDEMGFALMAHHEIAVYIEPIYOGAG GMRVYHEMLKRIKICDREGILLADELATGGRGKLFACHEAVIADIIICLKAL TGGTMTLSATLTREVAETISNGEACFMHGPTFMGNPLACAANAASIALLESQMOQ OVADIEVOLREOLAPARDAEMVADVILGALGVETTHVNNAAIÖKFEVEQGVIRP RGLTIYMPRYIILLPOOLRL/AAVNRVAVDDETFECQ"

gene
 CDS

gene
 CDS

BASE COUNT
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catgccagcattcgat 25
 ||||||||||||||||
 Db 2053 CATGCCAGCATTCGAT 2070

RESULT 18
 G01446 479 bp DNA SRS 19-AUG-1999
 LOCUS Dm0466 Drosophila P1 library Drosophila melanogaster SRS genomic
 DEFINITION clone DS07967 SP6, sequence tagged site.
 ACCESSION G01446
 VERSION G01446.1 GI:684849
 KEYWORDS SRS
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 RUBIN, G. (bases 1 to 479)
 RUBIN, G. Drosophila Genome Project
 Berkeley
 Unpublished (1994)
 COMMENT
 Contact: Drosophila Genome Project
 Berkeley
 Primer A: CTCCTACTGGGACTTGCTA
 Primer B: CACCAAGGCATATCCGA
 SRS size: 192

Contact: Drosophila Genome Project
 Berkeley
 Primer A: CTCCTACTGGGACTTGCTA
 Primer B: CACCAAGGCATATCCGA
 SRS size: 192

PCR Profile:
 Annealing: 58 degrees C PCR Cycles: 30
 Protocol:
 Template: p1 library Pools
 Primer: 1 uM each
 dNTPs: 250 uM each
 Tag Poly: 0.05 units/u1
 Total Vol: 15 u1

Buffer:
 MgCl2: 1.5mM
 KCl: 50 mM
 Tris-HCl: 50 mM
 pH: 8.3
 Gelatin: .001 %

The p1 library has been distributed to 16 regional sites. A list of these sites is available from Flybase, via anonymous ftp to <ftp.bio.indiana.edu> in the file <flybase/allied-data/genome-projects/1bl/1BL.doc>.

FEATURES

SOURCE

Location/Qualifiers
 1. .479
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="DS07967"
 /note="Vector: PAD10Sact1; The p1 library was made by D. Smoller in D. Hartl's lab (see Smoller et al., Chromosome 100: 487). Chromosomal position was mapped by polytene chromosome in situ hybridization in the Hartl lab (see Hartl et al., PNAS 91: 6824). STS's were generated by sequencing the ends of the Drosophila insert in these p1 clones, by the W. Kimmerly, C. Martin, and M. Palazzolo lab at LBL."

STS
 primer_bind 113 a 120 c 97 g 145 t 4 others
 BASE COUNT complement(233. .250)
 ORIGIN

Query Match 68.0%; Score 17; DB 54; Length 479;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcatgccagcatctcg 23
 ||||||||||||||||
 Db 396 GCATGCCAGCATTTTCG 412

RESULT 19
 AC019940/c
 LOCUS
 DEFINITION AC019940 53067 bp DNA HTG 03-JAN-2000
 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
 pieces
 AC019940
 VERSION AC019940.1 GI:6664957
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 53067)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 This sequence was identified as CDW:10211463 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source Location/Qualifiers
 1. .53067
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /db_xref="taxon:7227"

BASE COUNT 14944 a 11879 c 11720 g 14524 t
 ORIGIN

Query Match 68.0%; Score 17; DB 65; Length 53067;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcatgccagcatctcg 23
 ||||||||||||||||
 Db 13629 GCATGCCAGCATTTTCG 13613

RESULT 20
 AC013431
 LOCUS
 DEFINITION AC013431 153870 bp DNA HTG 31-JAN-2000
 Drosophila melanogaster chromosome X clone BACR22h11 (D1191)
 RPCI-98.22.H.11 map 13E-13F strain Y; cn bw sp, *** SEQUENCING IN
 PROGRESS ***; 73 unordered pieces.

ACCESSION AC013431.8 GI:6838815
 VERSION
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 153870)
 Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
 Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,
 Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,
 Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
 Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
 Sequencing of Drosophila melanogaster
 Unpublished
 2 (bases 1 to 153870)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT
 Submitted (11-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Jan 31, 2000 this sequence version replaced gi:6532028.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 73 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 995: contig of 995 bp in length
 * 996 1075: gap of unknown length
 * 1076 1560: contig of 485 bp in length
 * 1561 1640: gap of unknown length
 * 1641 2230: contig of 590 bp in length


```

* 2231 2310: gap of unknown length
* 2311 2956: contig of 646 bp in length
* 2957 3036: gap of unknown length
* 3037 3621: contig of 585 bp in length
* 3622 3701: gap of unknown length
* 3702 4535: contig of 834 bp in length
* 4536 4615: gap of unknown length
* 4616 5559: contig of 944 bp in length
* 5560 5639: gap of unknown length
* 5640 6615: contig of 896 bp in length
* 6616 7660: contig of 1045 bp in length
* 7661 7740: gap of unknown length
* 7741 8730: contig of 990 bp in length
* 8731 8810: gap of unknown length
* 8811 9404: contig of 594 bp in length
* 9405 9484: gap of unknown length
* 9485 10725: contig of 1241 bp in length
* 10726 10805: gap of unknown length
* 10806 11578: contig of 773 bp in length
* 11579 11658: gap of unknown length
* 11659 12834: contig of 1176 bp in length
* 12835 12914: gap of unknown length
* 12915 13501: contig of 587 bp in length
* 13502 13581: gap of unknown length
* 13582 14462: contig of 881 bp in length
* 14463 15566: contig of 1024 bp in length
* 15567 15646: gap of unknown length
* 15647 16366: contig of 720 bp in length
* 16367 16446: gap of unknown length
* 16447 17301: contig of 855 bp in length
* 17302 17381: gap of unknown length
* 17382 18390: contig of 1009 bp in length
* 18391 18470: gap of unknown length
* 18471 19409: contig of 939 bp in length
* 19410 19489: gap of unknown length
* 19490 20726: contig of 1237 bp in length
* 20727 20806: gap of unknown length
* 20807 22052: contig of 1246 bp in length
* 22053 22132: gap of unknown length
* 22133 23015: contig of 883 bp in length
* 23016 23095: gap of unknown length
* 23096 24622: contig of 1527 bp in length
* 24623 24702: gap of unknown length
* 24703 26689: contig of 1987 bp in length
* 26690 26769: gap of unknown length
* 26770 27825: contig of 1056 bp in length
* 27826 27905: gap of unknown length
* 27906 29904: contig of 1999 bp in length
* 29905 29984: gap of unknown length
* 29985 31588: contig of 1604 bp in length
* 31589 31668: gap of unknown length
* 31669 33112: contig of 1444 bp in length
* 33113 33192: gap of unknown length
* 33193 35374: contig of 2182 bp in length
* 35375 35454: gap of unknown length
* 35455 37223: contig of 1769 bp in length
* 37224 37303: gap of unknown length
* 37304 38978: contig of 1675 bp in length
* 38979 39058: gap of unknown length
* 39059 41337: contig of 2279 bp in length
* 41338 41417: gap of unknown length
* 41418 42461: contig of 1044 bp in length
* 42462 42541: gap of unknown length
* 42542 45823: contig of 3282 bp in length
* 45824 45903: gap of unknown length
* 45904 47247: contig of 1344 bp in length
* 47248 47327: gap of unknown length
* 47328 50153: contig of 2826 bp in length
* 50154 50233: gap of unknown length
* 50234 52182: contig of 1949 bp in length
* 52183 52262: gap of unknown length

```

FEATURES

source

```

* 52263 56756: contig of 4494 bp in length
* 56757 56836: gap of unknown length
* 56837 63525: contig of 6689 bp in length
* 63526 63605: gap of unknown length
* 63606 71274: contig of 7669 bp in length
* 71275 71354: gap of unknown length
* 71355 78141: contig of 6787 bp in length
* 78142 78221: gap of unknown length
* 78222 86570: contig of 8349 bp in length
* 86571 86550: gap of unknown length
* 86551 96227: contig of 9577 bp in length
* 96228 96207: gap of unknown length
* 96208 107432: contig of 11124 bp in length
* 107433 107431: gap of unknown length
* 107432 121618: contig of 14107 bp in length
* 121619 121698: gap of unknown length
* 121699 139797: contig of 18099 bp in length
* 139798 139877: gap of unknown length
* 139878 140388: contig of 511 bp in length
* 140389 140469: gap of unknown length
* 140470 141011: contig of 543 bp in length
* 141012 141091: gap of unknown length
* 141092 141513: contig of 422 bp in length
* 141514 141593: gap of unknown length
* 141594 142057: contig of 464 bp in length
* 142058 142137: gap of unknown length
* 142138 142690: contig of 553 bp in length
* 142691 142770: gap of unknown length
* 142771 143344: contig of 574 bp in length
* 143345 143424: gap of unknown length
* 143425 143857: contig of 433 bp in length
* 143858 143937: gap of unknown length
* 143938 144603: contig of 666 bp in length
* 144604 144683: gap of unknown length
* 144684 145116: contig of 433 bp in length
* 145117 145196: gap of unknown length
* 145197 145734: contig of 538 bp in length
* 145735 145814: gap of unknown length
* 145815 146224: contig of 410 bp in length
* 146225 146304: gap of unknown length
* 146305 146859: contig of 555 bp in length
* 146860 146939: gap of unknown length
* 146940 147585: contig of 646 bp in length
* 147586 147665: gap of unknown length
* 147666 148081: contig of 416 bp in length
* 148082 148161: gap of unknown length
* 148162 148539: contig of 378 bp in length
* 148540 148619: gap of unknown length
* 148620 148950: contig of 331 bp in length
* 148951 149030: gap of unknown length
* 149031 149375: contig of 345 bp in length
* 149376 149455: gap of unknown length
* 149456 149833: contig of 398 bp in length
* 149834 149933: gap of unknown length
* 149934 150532: contig of 599 bp in length
* 150533 150612: gap of unknown length
* 150613 151147: contig of 535 bp in length
* 151148 151227: gap of unknown length
* 151228 151577: contig of 350 bp in length
* 151578 151657: gap of unknown length
* 151658 151964: contig of 307 bp in length
* 151965 152044: gap of unknown length
* 152045 152629: contig of 585 bp in length
* 152630 152709: gap of unknown length
* 152710 153245: contig of 536 bp in length
* 153246 153326: gap of unknown length
* 153327 153870: contig of 545 bp in length.

```

Location/Qualifiers

1..153870

/organism="Drosophila melanogaster"

/strain="y: cn bw sp"

/db_xref="taxon:7227"

/chromosome="X"

Query Match 68.0%; Score 17; DB 63; Length 153870;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 7 gcatgccagcatttcg 23
 |||||||||||||||
 Db 63673 GCATGCCAGCATTTCG 63689

RESULT 21
 AC037435 183937 bp DNA HTG 15-JAN-2001
 LOCUS Mus musculus chromosome 11 clone RP23-9813 map 11, WORKING DRAFT
 DEFINITION
 AC037435
 AC037435.2 GI:12229535
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 183937)
 Birren, B., Linton, J., Nusbaum, C. and Lander, E.
 Mus musculus chromosome 11, clone RP23-9813
 2 (bases 1 to 183937)
 Anderson, S., Baldwin, J., Barnes, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
 Campione, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Melidim, T., Menus, L., Miho, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Rhmann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 15, 2001 this sequence version replaced gi:17528163.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 18882
 Center clone name: 98_T3

Summary Statistics
 Sequencing vector: M13; M77815; 4% of reads
 Sequencing vector: Plasmid; n/a; 96% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 170056 bases at least Q40
 Consensus quality: 177644 bases at least Q30
 Consensus quality: 180041 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 181637; sum-of-ctrls

Quality coverage: 4.7 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 base.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 23881: contig of 23881 bp in length
 23882 23981: gap of 100 bp
 23982 24469: contig of 488 bp in length
 24470 24569: gap of 100 bp
 24570 24755: contig of 186 bp in length
 24756 24855: gap of 100 bp
 24856 25997: contig of 1142 bp in length
 25998 26097: gap of 100 bp
 26098 26764: contig of 667 bp in length
 26765 26864: gap of 100 bp
 26865 28426: contig of 1562 bp in length
 28427 28526: gap of 100 bp
 28527 30488: contig of 1962 bp in length
 30489 30588: gap of 100 bp
 30589 32514: contig of 1926 bp in length
 32515 32614: gap of 100 bp
 32615 36175: contig of 3561 bp in length
 36176 36275: gap of 100 bp
 36276 38734: contig of 2459 bp in length
 38735 38834: gap of 100 bp
 38835 40927: contig of 2093 bp in length
 40928 41027: gap of 100 bp
 41028 43453: contig of 2426 bp in length
 43454 43553: gap of 100 bp
 43554 46744: contig of 3191 bp in length
 46745 46844: gap of 100 bp
 46845 50369: contig of 3525 bp in length
 50370 50469: gap of 100 bp
 50470 56799: contig of 6330 bp in length
 56800 56899: gap of 100 bp
 56900 57617: contig of 718 bp in length
 57618 57717: gap of 100 bp
 57718 68517: contig of 10800 bp in length
 68518 68617: gap of 100 bp
 68618 82793: contig of 14176 bp in length
 82794 82893: gap of 100 bp
 82894 96371: contig of 13478 bp in length
 96372 96471: gap of 100 bp
 96472 112208: contig of 15737 bp in length
 112209 112308: gap of 100 bp
 112309 127004: contig of 14696 bp in length
 127005 127104: gap of 100 bp
 127105 147474: contig of 20370 bp in length
 147475 147574: gap of 100 bp
 147575 174126: contig of 26552 bp in length
 174127 174226: gap of 100 bp
 174227 183937: contig of 9711 bp in length.

FEATURES

source

misc_feature
 /clone_11b="RP23-9813"
 1..23881
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 23982..24469
 /note="assembly_fragment"
 24570..24755
 /note="assembly_fragment"

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

George, R.A., Lewis, S.E., Richards, S., Ashnour, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Vandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazey, R.G., Chame, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabbor Miklos, G.L., Abtill, J.F., Agbayan, A., An, H.J., Basu, A., Andrews-Plamkoc, C., Baldwin, D., Ballew, R.M., Beeson, K.Y., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Besson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borikova, D., Borchan, M.R., Bouck, J., Brokstein, P., Brotlier, P., Butts, K.C., Busam, D.A., Butler, S., Cadieu, E.C., Center, A., Chandra, I., Cherry, J.M., Cartley, S., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Durov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferrara, C., Ferritera, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodok, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.T., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibeagwan, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Klip, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Mekullov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshireli, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacled, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J.D., Puri, V., Reese, M.G., Rehnert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Klamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spridling, A.C., Stapleton, M., Stroup, R., Sun, E., Svitskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, D.C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

2 (phases 1 to 327446)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7293108.
Location/Qualifiers
1..327446

complement(join(1/2..503,506..1504,1512..2077,2153..2813))
 4145)..>4380)"
 /gene="CG9281"
 /product="CT26402"
 /db_xref="FLYBASE:FBgn0009281"
 /db_xref="FLYBASE:FBgn0030672"
 complement(772..>4380)
 /gene="CG9281"
 /map="13E12-13E14"
 /db_xref="FLYBASE:FBgn0009281"
 /db_xref="FLYBASE:FBgn0030672"
 complement(join(772..903,968..1504,1572..2077,2153..2813))
 /gene="CG9281"
 /note="CG9281 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBgn0009281"
 /db_xref="FLYBASE:FBgn0030672"
 /protein_id="AAFA8493.1"
 /db_xref="gi:1293109"
 /translation="MSSAAKKRKAQKKKADRNKKIQDIPSTKKNVCPPELTLEKKL
 GKLEERKISAEARSCGSLAVHPRSRSDVLIANSTFFGSELDQTHLELNCGRY
 GLIGNGCGSLAVLIGREVPPVPIIDIFHLTRIPASSKALQCVMEVDERRIKL
 EKLAEELAMSEDDAQEQOLIDYIERLDDMSADLAEVKRAIILHGLGDKAMQOKAND

FSCGMRRLALRALFYKPHLLIDEPNTNHLIDACVLEBELKTYRILVLISHSD
 FLNGCTNLIHLTGKRLKYTYGNTAFAVTRMELLENOMKOYNEODIISHKNYIAR
 FGHGSAKLARAOSEKTLAKMAVAGLITEXVSDKVLNFPFSCGKVPPIYMONVN
 FRIIDETPWYIKNLIEGIDIDTLRALVOPNAGSKSTLLKILYGLVPTSGIKRNSH
 RTARYHOHLIELLDLDASPLEYMMRAFEDYKREEMKIRIGRGLTGOOVCPTROI
 DGORRVFAWILAMOVPHLLIDEPNTNHLIDETIDALADAINDFGSGMVLVSHDFRLI
 NOVAEIWECEKEVETWKMGSLIDYKDHKKKLTSEBKKAKAKK"
 join(4963, 5295, 5348, .>5848)
 /gene="CG15601"
 /product="CT35724"
 /db_xref="FLYBASE:FBan0015601"
 /db_xref="FLYBASE:FBgn0030673"
 /evidence=not_experimental
 <4963, .>5848
 /gene="CG15601"
 /map="13E14-13E14"
 /db_xref="FLYBASE:FBan0015601"
 /db_xref="FLYBASE:FBgn0030673"
 /evidence=not_experimental
 join(4963, 5295, 5348, .5848)
 /gene="CG15601" gene product"
 /note="CG15601"
 /codon_start=1
 /db_xref="FLYBASE:FBan0015601"
 /db_xref="FLYBASE:FBgn0030673"
 /evidence=not_experimental
 /protein_id="AAE48494.1"
 /translation="MSKLSPTKIPINFEAVRORCLYNTLIDSKKNVSEAYCA
 IIRSIKIPOTVSDIKTKISVTVYSKELRTIMRKELGRYERKLEFRIADSLR
 SVSLICKROGKNNSSAOITTKSDSTKLCTAADAITMSDALAEEDAVNGEPE
 ECLPESRPASTICKDSTLCLADPOEYHSGSCSSOOLPHITMAORSKYITSDS
 AGEDDILIFGQSIASOLRTIPDSYRSKVKLRIOQLPEAFEGFOSTRVNSTOJNT
 F"
 join(7954, 8220, 8556, .8795, 8865, .9086, 9145, .9416,
 9497, .9993, 10131, .10515, 10567, .15238, 15758, .18218,
 18282, .18755, 18814, .19154, 19217, .20287, 20378, .20513,
 20573, .22032, 22060, .23197, 23270, .23671, 23739, .23927,
 23992, .24313, 24377, .24542, 24617, .24807, 24885, .>24987)
 /gene="CG8184"
 /product="CT24358"
 /db_xref="FLYBASE:FBan0008184"
 /db_xref="FLYBASE:FBgn0030674"
 <7954, .>24987
 /gene="CG8184"
 /map="13E14-13E16"
 /db_xref="FLYBASE:FBan0008184"
 /db_xref="FLYBASE:FBgn0030674"
 join(7954, 8220, 8556, .8795, 8865, .9086, 9145, .9416,
 9497, .9993, 10131, .10515, 10567, .15238, 15758, .18218,
 18282, .18755, 18814, .19154, 19217, .20287, 20378, .20513,
 20573, .22032, 22060, .23197, 23270, .23671, 23739, .23927,
 23992, .24313, 24377, .24542, 24617, .24807, 24885, .24987)
 /gene="CG8184" gene product"
 /note="CG8184"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008184"
 /db_xref="FLYBASE:FBgn0030674"
 /protein_id="AAE48495.1"
 /db_xref="GI:7293111"
 /translation="MKVDYVILKKGSELTAECEHLLIDELTKRRAELFLDITVY
 WYKCELIYHIEILDFEDKILFEASRVHNGNEVLCDDNFOFTDTLLHNTYITV
 LILBFSRRLYNSIEHLVLLSSONMDIYAVILNLYMSKSNFIPLRPFERELL
 IVKLFTAEKRGDNGVLSIKDCDICEPKLEFLYDCTIDVDEGHAAQLEIPDMDL
 CHTASAEVITKTIAGQISKESEATKRIIAHVRRLISGFNNYKLRLOFVQARLAVSTL
 IYSNALDNTDKVLYAGFEGELCELIDKEOVHVLRTLRAVALRTLSMLHFRDNPVS
 BASRKLKIVHYTGAEROGGTLPLLVDCIDINDITBGLTRRYPLTATSLEFLIHA
 SYEKGATLAKVSGMMOSLCLVSIWPGVDLHITFVPAVVIDLITNIDIAHONG
 LNVFIDILKEIKSCCALADIGITITKESTLRDNDKIDTSLIDVDEDEVEGSENNASV
 TSGNASPRDNDSDGATGAGSIYRAYHEHICGAGSAGLIDIGARSSLSGSSGTS
 GMGAGSINRDSASARPAIMPTNYSYARLARKEELSSQLPSSLAPKPPSCVTOGA
 ALLNNMNTMETSIJOSLRHITANAEEYGSPLFLATDVVTYVFNPEPSLISLDIGA
 ITSVMLKLLQKQVPAITREVLGSI,PNVFSALCLNENGLFEFLISDPEDKVLKVLSPD

Query Match
 Best Local Similarity 100.0%; Pred. No. 8.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 7 gcatgccagcattccg 23
 Db 70808 GCATGCCAGCATTTCG 70824
 RESULT 23
 AF146614/c
 LOCUS
 DEFINITION
 DdbC precursor (dbsc) gene, complete cds; and RecJ (recJ) gene,
 partial cds.
 accession
 version
 keywords
 source
 organism
 AF146614 1186 bp DNA BCT 09-AUG-1999
 Pectobacterium carotovorum subsp. carotovorum.
 Pectobacterium carotovorum subsp. carotovorum
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Pectobacterium.
 1 (bases 1 to 1186)
 Vincent-Sealy L.V., Thomas J.D., Commander P. and Salmon G.P.
 Erwinia carotovora DsbA mutants: evidence for a periplasmic G-stress
 signal transduction system affecting transcription of genes
 encoding secreted proteins
 Microbiology 145 (Pt 8), 1945-1958 (1999)
 MEDLINE 99392457
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 Thomas J.D.
 Direct Submission

YLVAAMRRSSDPLGDTATNIGNANDI KHHPTLRADATEA IVLLNELVRLGSDPS
 FLOWRANKSSSGSGSGSGHGGHHVATPSPMVAVAGSGASVSLQGAADNSGSD
 DDDDDDEMSASQ000PTTPGOGGSPTRPVQAGGVGAGATGPAADNSQAVK
 VTPPEEAEPLIDYILINWKAIEALFVSNPNKDORREVLHGLKPIITLQSLPLPV
 DSVATTSOAVANVCALLSQOAEFRVLDAVNOKLADIVAOGLKPIIKHETPGGSLV
 AELVCCORLEDFANAEYPTILHNSNFGYVVMVHLICRANSMRTIILKRCVNR
 ETGVLOOLOOLYISLWNESTILNCSDEPTORHDLIMPEIAQMSAAAGTADPLT
 EAEISKLEIRAAEFRTKYPPEORFYSILASRRRLRALLGLITLVKLSVSPQT
 RORINDILSNITKYPPEARIATILSLINWGHSHSITLKPVPKLTFLICSVG
 PTGPMFLFDRKSAFLHMTSOPCAENGLEAFEMFWALSDLNVSQKFEEDWASMEDL
 MOSHEDKDCPEGTEGEFDLAWOLLEKVNPRAMIDSPNPSPLINWSPVPEPV
 FYLIHQLAMOLRIRIARREIPNYGSMETMILLIKLIKSHQTLDDYRPMKE
 IKTENLYIKSTDDGLNADHIKITLDMGNHVIYALITNLSAIEATYDILNPEAS
 LSTTGOSSSAPPPPPSASTMDIDVDVADGDSQKSTSPNSDYDKHLKLP
 SLIFDFOEAISRAPEFMEALPDVTNAAOLIVLYIRRHNLKQVFTVFNNTIO
 WYDFTLOLPEPAKTTGSKASRLBECLTMTATRLRVLKSTLLLEENY SOLHRLVE
 ATTADAVSLVKILDCMSOMLEOSTETNGQOAVPVPRVQNLVYDLIDAIDSISHI
 LORANMRVSSDVWRMYDASTGKNVASEANNELIRNAVAGGRMILYINRGRCY
 SLNCTIOVSEASGTHRPVCPALIKLEAITSNNPVALHKEVHILNRYRTAPDGSY
 HSDDELISRLQVNFIDGQPOOSENSANPSEATEEDGPAANGSPATMTTSKR
 OMSAAAAAANKKSQSGSPPIPKRTOKIILNEOYGLSKDCCRTIGSYDILQPSH
 VLEAPPVLOEAMERYLGMRAAGITYPVGRHDLITVLOVSSAVSNPQI EVAAKEMLK
 GEYLINTSTTLADDAVAVAKSKFGQOPTAAAGATPHNQPPTPPAPAKELKDDPV
 QSSVAVLELLHGLVOPCMHAGYGSILDRGGAELLQPDQPTGTAAGCHDDEMTVLY
 ICSLSDCYOOPGOVHVAEELHAAVARALAPDSEAKHRTLOVMSLEPNLIESPMY
 DVIHNRNNMYRVLIRGVMTYITKVAOYKRLSHNLTSLAATIRHETLIRPSVT
 TTIGSKRLIFPSSGPGPNASGNNINNAAGSGSGPSGNTGATIRRLYPLAAAT
 GSGSAMERSNAMGEBHYLRIIRVLDRIHAPFTINSDMAPYSDDAAPAGSGVG
 GPVPVYNEPDLMDPFIQSBELRLASAGGAGSGSVSPANVSSASGAGOSTSG
 SGNNSGICNLKRLNPVDVDSGSGSGSGSNAGVYNDGIISGGSIDASSTSDG
 AEGGARDAVLLGDVSTESDPSSTENENEDDEDDADEVDDEHSTEDVE
 ERSQFTEVPHIPEPSEETAADAVDDDDDEDVDDDDDDDDDDDDDDDDDD
 SYMAEIERIDIALVSSNNRRRAAVANASDGSANSTNDPDDDDDDDDDDDD
 DGDVDVPISGPSAQAELYERLIGSLMPTPVPGRVYRNNSTIAOPLDITVNGNA
 GSGSNACATSGSGSGSGSGSGSTAGTATGATASALRPSGSMAPDFNTRISGDS
 SMGSGPAICGANAGDMNVITPMSLERRGAGSGVOTNELNPDQACGAGGSGG
 SASGVGAVGAGGTGGHPNATCTACTVNAHRLYLIDQYCVCMPTNQPTEYMTF
 TQDLHWMLEAVTLIDMESQTDACIVAAHLLPHLLIKELKAAHQKREBARANSTTP

JOURNAL Submitted (29-APR-1999) Biological Sciences, University of Warwick,
Gibbet Hill Road, Coventry CV4 7AL, UK

FEATURES
Source
1. 1186
/organism="Pectobacterium carotovorum subsp. carotovorum"
/strain="SCRI193"
/sub_species="Carotovora"
/db_xref="taxon:555"
<1. 189
/gene="xerD"
/gene="xerD"
/note="similar to Escherichia coli xerD
integrase/recombinase in SwissProt Accession Number
P21891"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="XerD"
/protein_id="A047611.1"
/db_xref="GI:5712697"
/translation="KLSPLVLRHAFATHLINHGAGFTGRTDALGAQRSEFDADLHPCR
DGAETRASAAHVRHRT"
222. 1031
/gene="dsbc"
222. 227
/gene="dsbc"
246. 251
/gene="dsbc"
258. 264
/gene="dsbc"
/note="putative"
315. 377
/gene="dsbc"
/note="putative sec-dependent export signal"
315. 1031
/gene="dsbc"
/function="periplasmic disulfide bond formation"
/note="disulfide isomerase; similar to Erwinia
chrysanthemi and Escherichia coli dsbc proteins in
SwissProt Accession Numbers P39691 and P21892"
/codon_start=1
/transl_table=11
/product="dsbc precursor"
/protein_id="A047610.1"
/db_xref="GI:5712696"
/translation="MKKGLLLSLVATATNFADDAIAIKQTILASGYAGGDPFA
DGRKVTITDSCVLYISEGQAFAPALRRERTMPNTNKLILKDLALQOMTV
KAAQEKVITVFSTIGCYCHKLHEQKMDYNALGITVRYLAFRQGMKSPAKDMQSI
MCVADRKAQPSAMKGDALSAATCKTDIAHYDLGIQGVQGPATLVLDGLVPGYQ
GPKMLAMLEHAKSKRTGC"
378. 1028
/gene="dsbc"
/product="dsbc"
669. 680
/gene="dsbc"
/note="region: putative CPHC motif typical of the active
sites of disulfide isomerases"
1044. >1186
/gene="recJ"
1044. >1186
/gene="recJ"
/note="theoretical protein: similar to Erwinia
chrysanthemi and Escherichia coli RecJ single-stranded
DNA-specific exonucleases in SwissProt Accession Numbers
P39693 and P21893"
/codon_start=1
/transl_table=11
/product="not experimental"
/evidence="not experimental"
/protein_id="A047612.1"
/db_xref="GI:5712698"
/translation="MITQLRRRLAEDIDLPTVPLLRLLRYAHGVKGAQQLERSIS

BASE COUNT 295 a 315 c 323 g 253 t
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 1186;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcacgcacagcattc 22
Db 993 GCATGCCACGACATTTC 978
RESULT 24
AK024397/c
LOCUS
DEFINITION
Homo sapiens cDNA FLJ14335 fis, clone PLACE4000411, highly similar
to Homo sapiens mRNA; cDNA DKFp586D0624 (from clone
DKFp586D0624).
ACCESSION
AK024397
VERSION
GI:10436775
KEYWORDS
Oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens placenta cDNA to mRNA, clone_11b:PLACE4
clone:PLACE4000411.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 4674)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5' - 3' end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source
Location/Qualifiers
1. 4674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="PLACE4000411"
/clone_11b="PLACE4"
/tissue_type="placenta"
/note="Cloning vector: pME18SFL3"
608. 1273
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14908.1"
/db_xref="GI:10436776"
/translation="MPSIKYLGWTLVEOPKGEELSAAIKRIYATKASGKRLQKVT
LKVSRRGILLDNTLNLIENVSTYCTADKMBKVPAYIAOSHOHSLCEHAF
CTKRMAQAVTLTVAAQKVAEPFQWQSKKEKMDKASQEGDIVGARDGCPPKS
LVATGNLLDEETAKAPLSTVSANTTNMDEVPRLQALSGSSVWVSCVARSVILSL
TSG"

BASE COUNT 991 a 1342 c 1239 g 1102 t
ORIGIN

Query Match 64.0%; Score 16; DB 89; Length 4674;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgcatgcgccagcat 19
 |||
 Db 2446 ggtgcatgcgccagcat 2431

RESULT 25

AF206717.c 11910 bp DNA SYN 12-SEP-2000
 LOCUS AF206717 Shuttle vector p13 hypothetical proteins, HtrU (htru), Ardu (ardu),
 DEFINITION RepU (repu), and Resu (resu) genes, complete cds.
 ACCESSION AF206717
 VERSION AF206717.1 GI:7262998
 KEYWORDS

SOURCE

ORGANISM Shuttle vector p13.
 SHUTTLE VECTOR p13
 artificial sequence: vectors.
 1 (bases 1 to 11910)

REFERENCE

Meima, R. and Lidstrom, M.E.
 Characterization of the minimal replicon of a cryptic Deinococcus
 radiodurans SARK plasmid and development of versatile Escherichia
 coli-D. radiodurans shuttle vectors
 Appl. Environ. Microbiol. 66 (9), 3856-3867 (2000)

JOURNAL

MEDLINE 20422197
 PUBMED 10966401
 REFERENCE 2 (bases 1 to 11910)
 AUTHORS Meima, R. and Lidstrom, M.E.
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-1999) Chemical Engineering, University of
 Washington, Benson Hall, PO Box 351750, Seattle, WA 98195, USA

FEATURES

source location/Qualifiers
 1..11910
 /organism="Shuttle vector p13"
 /specific_host="Deinococcus radiodurans SARK"
 /db_xref="taxon:115340"
 /note="derived from pUE10"
 105..593
 /note="orfA"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44052.1"
 /db_xref="GI:7263004"
 /translation="MGQIVBGGDDVPPNDGQGVVHRLPLFVGVKDKFAMLRG
 GKRONPDGDISWPKRMVVEAFHPDDGGGVRAGRGSEFSPMDVLAPEGDEHGG
 EQLDLVLAELREVREAGAGOLGQDIGRVPFSKSRTPSPAGNCRVLRMGQKRPK
 AV"

CDS

misc-feature

complement(257..683)
 /note="5 mismatches with sequence deposited in GenBank
 Accession Number M94966"
 649..653
 659..1078
 /note="OrfB"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44047.1"
 /db_xref="GI:7262999"
 /translation="MREVVAFDGSSEPSQALKAQITLGAAGAELEVVTVEPPVSG
 FVQWAGLNPVYREMLEQNAQOQTOTLIGFVGQVGRVOTLGRIRLEVLTEAHAV
 DLVVGTHGYRGMDRIILGSVTEFLTRAEVPLLVYR"

gene

1187..2436
 /gene="htru"
 1187..1192
 /gene="htru"
 1198..2436
 /gene="htru"
 /db_xref="GI:7263001"

RBS

CDS

/codon_start=1
 /transl_table=11
 /product="Htru"

/protein_id="AAF44048.1"
 /db_xref="GI:7263000"

/translation="MKTAMPILTLILLACGTDPTESRTQTTSRTATNPAFGA
 ARDSSPSGPAQGTASPDQRLDYDQNTIDVVKTDQGVVFTREDDGGGALFSESP
 LAPDAPEDEQEGTSGSGFLIDREGHITITVHYVODATDRIIRHOMERXYPASVY
 TAPADLALIRASGARRPAGAEAGGRRGRRGDDTGAIPGLFTVTGIIIPAYK
 RYIPMGVESIPONSQVTDATIPNGSGCPLYNSRGEVVGNTQILSPAGAVTGQNA
 GVGFAIPVNAVKSLLPLRAGETITPRTIGVSVNLQALTPSAREALGLPQGVLYVS
 VEGTPAAAGKRGGRSQREPDGDIRLGSDIITAVDQEVSTVODLQGVLLGKQSGD
 EVIILLRDQGTILORRKLTS"
 3113..3118
 3127..3678
 /note="orfC; proline-rich protein"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44053.1"
 /db_xref="GI:7263005"

RBS

CDS

/translation="MTHCAAMASARSTRSQRCAVTHRSALGCVGESAAVREPAVY
 ARLPVPPVTSAAPEFGHLDSDMHRPAPSAALPPVPPVPSRWGNCSPAVPLSRIP
 RPSSTPHIKISFCWGRCSASPLILPKQITVPGRAAGGPPILRPHFWPARPACFERCP
 AARPLVSLSCIVCAVYLLSCI"
 3691..4038
 /note="orfD"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44054.1"
 /db_xref="GI:7263006"

/translation="MTTAPLFSASAVISRPQSVNVRSGCCAAAMWCKNRRTAP
 ALAIRPASAAPSTTAATVARMYQNFGLRVQRRAPCCVYLKSGPPAPDLAAQ
 WMAIVAYAGGRQ"
 4090..4531
 /note="OrfE"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44055.1"
 /db_xref="GI:7263007"

/translation="MAAPTARTHSPPMLPLAGLMSRVAVALLTPPLPPPCSAV
 TPMPACACFVPLVTPPTSPVPGRRSPPPPSPCFPPRGSAPARRMPCGWPSSAA
 CPCESLAPLRPPRPARGRRPPLQAPFAGSGRRRRCPACSSFPAPARSFL"
 4666..4670
 4678..5364
 /note="OrfF"
 /codon_start=1
 /transl_table=11
 /evidence="not experimental"
 /product="hypothetical protein"
 /protein_id="AAF44056.1"
 /db_xref="GI:7263008"

/translation="MSELITFRISPDHARALVALGADHTAPALARAADLVMOQISGP
 APVTDAIRRSAPARLRLPLGVGYGWDSDAPLIPARACKLNLTLTGAADPTGMAR
 ALTGNDITAEQITCEGMDLIDIONHRRAOIAAQQOQAPAPAPLITRTQARALHK
 QIAGIVTEDEERYTQAAKIARRPVESITLALNODEAALLVVAABEABRBAVLAQD
 PAAYGALLR"
 5407..5973
 /gene="ardu"
 5407..5412
 /gene="ardu"
 5422..5973
 /gene="ardu"
 /codon_start=1
 /transl_table=11
 /product="ardu"
 /protein_id="AAF44049.1"
 /db_xref="GI:7263001"

RBS

CDS

gene

1187..2436
 /gene="htru"
 1187..1192
 /gene="htru"
 1198..2436
 /gene="htru"
 /db_xref="GI:7263001"

RBS

CDS

RBS		/translation="MTYTHPLIERHPDAPLALMGCLAAVNAACKLGAMMOKASSTPAE MGALPEITAKSPERPHAEEMIDIMDDMMPPTEAGPTLSATYTAALGLADSPADAEEV AAMVEKGAEEKMDADKTDTDAYTIGRPDSVEDTIAOYLDDSGALOEVBEMLRPIYINFAAL GRDMELNDGVYEGKNHFFNGHA"
RBS	6148 . 6153	
CDS	6164 . 6445	/note="orf"
		/codon_start=1
		/transl_table=11
		/evidence=not_experimental
		/product="hypothetical protein"
		/protein_id="AAF44057.1"
		/db_xref="GI:7263009"
		/translation="MOVARMDKPILPPEVAEDVLTVKWLWLYTMANPGCHSDRSILMA ELGVRGOAIKRLLRGVLYVOEAPRGPKPKRYAQKKAKKPAOEENP"
RBS	6736 . 6741	
CDS	6748 . 7749	/note="OrfH; alanine-rich protein"
		/codon_start=1
		/transl_table=11
		/evidence=not_experimental
		/product="hypothetical protein"
		/protein_id="AAF44058.1"
		/db_xref="GI:7263010"
		/translation="MOKLTAAAILFSAIALTAGCGPNTTPQAPAFSAVDPTDG ARVLASTGRMSAAQAVOSLIGRLDLATAAPATVTYVAVMGTGPSRAVGORRALIG WKISGCPFAHLTGATSGSYALALTDPGADVKTGTGLSVSLPLASGNINTSRSSST KYAAVGTGCVAHSALLTLANGSAAEFCTGAAPLVVCAGAGYAPVPTNPAPATLELD MWSNFTTSAGRFAGLIVFAQVPDKAWMLTKYOYQAEECCAGDAVAAAGASEMOT LSSPANTRRPFVAVDEKMKVQEHYGLAYQFAFKDEGPTRHGCTLLREBLDFPGATFS" 7953 . 9159
gene		/gene="repU"
RBS	7953 . 7957	
		/gene="repU"
CDS	7966 . 9159	
		/gene="repU"
		/function="Initiation of plasmid replication"
		/codon_start=1
		/transl_table=11
		/product="RepU"
		/protein_id="AAF44050.1"
		/db_xref="GI:7263002"
		/translation="MRHNPEFWRIIEAGLIQSILQENHAKRPALTFREISSARSIS PPPYPTTAOSOPAPAPILTATRAPLNYOTLTACLDAPCRSARRIFRALPSVALDVA RACGYASHRTAVFHLPABELLMHVGLKSATFYEMLOTLRVGLYACDACHMGDIRES VATGTLAAVTLKPRVLEVKGAGTVRLMHDDWRNRMDLNADRKAGRTYYNMILHPLPS VPSEOCBEDCVTGGLEREKSAIPAYTYEVYSTWALKSALPSPSDMTYVRAPSIAENV VWEILDADHRETFRAYRAETIYDRQARALAAAFCDGADSLGFPMKLITWNITRAADAGRIV SDMGAVLVIRLDVKHKDQTAGCTPPRNLAAYVNALADLIGRLREYEGORGVPRDQ AOFAA"
protein_bind		complement(8717 . 8725)
		/note="putative"
		/bound_moiety="DNAa"
protein_bind		complement(8745 . 8753)
		/note="putative"
		/bound_moiety="DNAa"
misc_feature	9164 . 9637	
		/note="region of low GC; possibly involved in replication initiation"
RBS	10614 . 10619	
		/gene="resu"
gene	10614 . 11590	
		/gene="resu"
CDS	10625 . 11590	
		/gene="resu"
		/note="resolution of plasmid multimers - plasmid
Query Match	64.0%; Score 16; DB 56; Length 11910;	
Best Local Similarity	100.0%; Pred. No. 45;	
Matches	16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	8 catgccagcatctg 23	

Db	11110	CATCGCCAGCATTTGC	11095	
RESULT	26			
LOCUS	DMBH48C10			
DEFINITION	DMBH48C10	49587 bp	DNA	INV
ACCESSION	Drosophila	melanogaster	BAC	clone BACH48C10.
VERSION	AL133503			
KEYWORDS	AL133503.1	GI:6594136		
SOURCE	fruit fly.			
ORGANISM	Drosophila	melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 49587)			
AUTHORS	Murphy, L., Harris, D., and Barrell, B.			
TITLE	Unsequencing the distal X chromosome of Drosophila melanogaster			
JOURNAL	Unpublished			
REMARK	Sanger Centre, Wellcome Trust Genome Campus, Hinxton Hall, Hinxton,			
	Cambridge CB10 1SA, U.K.			
REFERENCE	2 (bases 1 to 49587)			
AUTHORS	Benos, P.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-DEC-1999)			
COMMENT	European Drosophila genome Sequencing			
	Consortium			
	Clone-BACH48C10; Contig ID=1; Length=49587; Status=Finished			
	db			

FEATURES
SOURCE

```

ftp://edgserver.rockefeller.edu/
The syntax for the representation of annotation used in this record
is documented at:
ftp://ftp.ebi.ac.uk/pub/databases/edg/sequence_annotation.README
Coding sequences are predicted from computer analysts, using both
gene and CDS prediction programs and matches to other sequences.
These predictions and matches have been evaluated by the annotators
and may have been refined by hand (in which case a Genefinder
prediction will have no score. The annotators have also used their
judgement on what matches to represent in this record. A far more
complete annotation record is available from FLYBASE
(http://flybase.bio.indiana.edu/) through the FlyBase Annotation
Object linked by the db_xref qualifier in the Feature Table.
IMPORTANT: This sequence is NOT necessarily the entire insert of
clone BACH48C10. It may be shorter, since we are minimising the
overlap between clones to 100 bases, by trimming them. Clone
BACH48C10 overlaps to the left with clone 82C7 The true left end of
clone BACH48C10 is unknown
Clone BACH48C10 overlaps to the right with clone BACH7M4 The true
right end of clone BACH48C10 is unknown
Sequence in absolute orientation with respect to chromosome.
location/Qualifiers
1.49587
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC BACH48C10"
complement(530..1394)
/gene="EG:BACH48C10.1"
complement(join(530..1009,1095..1394))
/gene="EG:BACH48C10.1"
/note="/prediction(method:"genefinder",
version:"084", score:"21.51")";
/prediction(method:"genscan", version:"1.0",
score:"73.25")"
/codon_start=1
/protein_id="CAB63507.1"
/db_xref="GI:6594143"
/translation="MSKKYDNPVTKHFFVSEPTRRSKCHFDMAGRHSNLRHILK
RKHEHDAYSLVDKREIRALAOQAKDAVNPRTKVLASKYPLYPPTRKOSFTTEDD
DSOHSGEIWRVTLNGEETGCDYDNKNNAVTSQFEVFLGKSRIPDEEYQWKNAMELK
ADVSAAGTTNNESSVCAPSTTARARSPRAASSAAALDOSPDEDTFLRYLQSKLGLG

```



```

gene
  CDS
    KYAARTVQFHNRLIKKADMGHFEETSRMCDDDYASP"
    2911..4557
    /gene="EG:BACH48C10.2"
    /complement(join(21238..24581,24649..24951,25046..25103))
    /gene="EG:BACH48C10.5"
    /note="/prediction-(method:"genscan", version:"1.0",
    score:"314.45"); /prediction-(method:"genefinder",
    version:"084"); /match-(desc:"G02860.5prime GH
    Drosophila melanogaster head port2 Drosophila melanogaster
    cDNA clone GH02860.5prime, mRNA sequence",
    species:"Drosophila melanogaster (fruit fly)",
    ranges:(query:25045..25290,
    target:EMBL:A1064644:369..124, score:"1221.00"),
    (query:24786..24952, target:EMBL:A1064644:535..369,
    score:"817.00"), method:"blastn", version:"1.4.9");
    /match-(desc:"Ld12603.5prime LD Drosophila melanogaster
    embryo Bluescript Drosophila melanogaster cDNA clone
    LD12603.5prime, mRNA sequence", species:"Drosophila
    melanogaster (fruit fly)", ranges:(query:25040..25317,
    target:EMBL:AA438345:305..28, score:"1363.00"),
    (query:24647..24952, target:EMBL:AA438345:604..299,
    score:"1422.00"), (query:24549..24579,
    target:EMBL:AA438345:635..605, score:"155.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"bs12f02.y1 Drosophila melanogaster adult
    testis library Drosophila melanogaster cDNA clone bs12f02
    5', mRNA sequence", species:"Drosophila melanogaster
    (fruit fly)", ranges:(query:22494..24011,
    target:EMBL:A1945413:522..5, score:"2554.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"GM04080.5prime GM Drosophila melanogaster
    ovary Bluescript Drosophila melanogaster cDNA clone
    GM04080.5prime, mRNA sequence", species:"Drosophila
    melanogaster (fruit fly)", ranges:(query:22290..22894,
    target:EMBL:AA802249:605..1, score:"2971.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"GH25306.5prime GH Drosophila melanogaster
    head port2 Drosophila melanogaster cDNA clone GH25306
    5prime, mRNA sequence", species:"Drosophila melanogaster
    (fruit fly)", ranges:(query:21634..22176,
    target:EMBL:A1405337:543..1, score:"2706.00"),
    method:"blastn", version:"1.4.9");
    /codon_start=1
    /protein_id="CAB72240.1"
    /db_xref="GI:6911911"
    /translation="MSSSESTGCDSDLYDPLAEELHNVOLVHRTRENIDALNAKEAN
    LQEPAMVLTIGESKALNTWVLTGNPTSKLEIKPTVYVYSRLMEFETFEOL
    TSLHLEAKQOEIMERLNSODOEDSLVERKPOPHYONOTOILOOOROLARVHHC
    NDLTDLGSDPSOGGTLROPKILRALPNOQRTSVESVIGRLCDLMALALRO
    LTPDMCEVSTSGRHIIIPMHDTGLHVEELFVRLLDEPIRTHKQITIKTFEEL
    VCEGCRRLITPFGSCOCNRRPHQCANRPMILQPPMDSYOLLAEENDNGVGF
    PGRTAVRFNMSKRSRSCSSSSSSSSSSSGNHRGPRPISODDSSNAPN
    VCTINIRVTSYEVORSLIQARPPLPHPTDSTNSTQASPTSLKHNRPRASADESN
    KILLRLDAKSENNNIIAEELILGPRISSGTVYRAHMGPAVATLVANKTPSPA
    QLOARNEVAAMLKTRHCNIIILPMGCVSPSLAIVOMCEGSLKKHVAHSPTEKLN
    TLIDIGROVAGMDYLAKNIIHRLKSNFIILHEDLSVKITDEBLATAKTRMSEKO
    ANOPGSLIMANAEVITRMDELNPYSFOSVYVAFGVIELLAECLEPGHISKDDIIF
    MYRCILRPDMQSVSDAPQALKRLAEDCITTPKDRPLFRLLMLNMLRTLPKIH
    RSASEPNLTOSOLONDEFLYLPSPKTPVNFNPFSGAGNT"
    /complement(19754..20239)
    /gene="EG:BACH48C10.6"
    /complement(19754..20239)
    /gene="EG:BACH48C10.6"
    /note="/prediction-(method:"genscan",
    version:"084", score:"26.59");
    /prediction-(method:"genscan", version:"1.0",
    score:"82.07"); /match-(desc:"SIMILARITY TO RIBOSOMAL
    PROTEIN L14", species:"Caenorhabditis elegans",
    ranges:(query:20027..20128, target:SPTREMBL:Q20430:36..3,
    score:"94.00"), (query:19952..20011,
    target:SPTREMBL:Q20430:67..46, score:"73.00"),
    (query:19835..19924, target:SPTREMBL:Q20430:109..80,
    score:"88.00"), method:"blastx", version:"1.4.9");
    /codon_start=1
    /protein_id="CAB63504.1"
    /db_xref="GI:6594140"
    /translation="MALRIITKRVQVAPLAVTIGCOOTQOLHTTACEITRLANL
    RVVNSDGKKAMAGRPRTCIHYVKNKGVEFTDKVLAITKGMKKGIIIVGLKONOK
    PKQKFTDSNNIIVLIDNGSPGTRIHVPIPIRLITLKEKTLAKGADYTKVLAIASRY
    V"
    /complement(21238..25103)
  gene

gene
  CDS
    /gene="EG:BACH48C10.5"
    /complement(join(21238..24581,24649..24951,25046..25103))
    /gene="EG:BACH48C10.5"
    /note="/prediction-(method:"genscan", version:"1.0",
    score:"314.45"); /prediction-(method:"genefinder",
    version:"084"); /match-(desc:"G02860.5prime GH
    Drosophila melanogaster head port2 Drosophila melanogaster
    cDNA clone GH02860.5prime, mRNA sequence",
    species:"Drosophila melanogaster (fruit fly)",
    ranges:(query:25045..25290,
    target:EMBL:A1064644:369..124, score:"1221.00"),
    (query:24786..24952, target:EMBL:A1064644:535..369,
    score:"817.00"), method:"blastn", version:"1.4.9");
    /match-(desc:"Ld12603.5prime LD Drosophila melanogaster
    embryo Bluescript Drosophila melanogaster cDNA clone
    LD12603.5prime, mRNA sequence", species:"Drosophila
    melanogaster (fruit fly)", ranges:(query:25040..25317,
    target:EMBL:AA438345:305..28, score:"1363.00"),
    (query:24647..24952, target:EMBL:AA438345:604..299,
    score:"1422.00"), (query:24549..24579,
    target:EMBL:AA438345:635..605, score:"155.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"bs12f02.y1 Drosophila melanogaster adult
    testis library Drosophila melanogaster cDNA clone bs12f02
    5', mRNA sequence", species:"Drosophila melanogaster
    (fruit fly)", ranges:(query:22494..24011,
    target:EMBL:A1945413:522..5, score:"2554.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"GM04080.5prime GM Drosophila melanogaster
    ovary Bluescript Drosophila melanogaster cDNA clone
    GM04080.5prime, mRNA sequence", species:"Drosophila
    melanogaster (fruit fly)", ranges:(query:22290..22894,
    target:EMBL:AA802249:605..1, score:"2971.00"),
    method:"blastn", version:"1.4.9");
    /match-(desc:"GH25306.5prime GH Drosophila melanogaster
    head port2 Drosophila melanogaster cDNA clone GH25306
    5prime, mRNA sequence", species:"Drosophila melanogaster
    (fruit fly)", ranges:(query:21634..22176,
    target:EMBL:A1405337:543..1, score:"2706.00"),
    method:"blastn", version:"1.4.9");
    /codon_start=1
    /protein_id="CAB72240.1"
    /db_xref="GI:6911911"
    /translation="MSSKMGSPATINMPFRKKDKSGTSSMGKLSKSTQJLHLST
    DTQINNSOLEFNRITTPVPAATSNAPPEQFRITVLDPKQJLVARLGAAPVLSKLS
    LVCDNKLDDKTEFRSPVDASQVYSESTIGVGSJELRLCKRSSYDNFNVDVHK
    FQRTVARDDLSKSSDSSSRHSKSVYAKTPSPYSSNSLMSDSTGMNTRFVVVPP
    AKVLAAPPKRTAPRPSQVCIPEGVDPDIPANAKSEPAVAVIKPOLCSTPNL
    SGPLELDGNGSGSPDESDTGHYATLDLKPITAGSEPTPKRLQIKKTPAPAP
    PDROSQDNTSLASLPEPPVPPNIPQAPRITTPLPAPATPAQVYNGNGSPN
    GNSKVLINRTPPEPRISGATEDDNDPQASQAGICEPAPSPSYPEPAESQ
    QESSDEDAIKYNNRFLCOTSKKPPESATITAEALTAADIDEEDNVDDPASTIN
    GNATPSSSETSLTSMNLTDPISPPDSDDQWRNAGSLSPSPDELTVDELATIN
    QNRDLTIKQPPQVWEIESAKPNRIANSLTASTRIGADSFQAGASGGGTEV
    ASSGALSHRSSSHSVSLNKLEONGICIGORSSSELSIESLSQEPVITKILNKRK
    NSLAERSGASSTVTEQPKITIKOLESIAKQSSVDSDIISRSRSPASQSPPVK
    SNPGVEKPTSPTRAVQOEPYSPSALOKSSPAVQORPTSPVPEPSIPVAVKVEBAY
    SLHOKELVYOKESPLESKESPVYALOKSSPAVYKPPPPKPPVPEVAKSSGNSNSO
    SITTPRTATPVPNALRPNMTMAETEOSQDEPAFSPKPLPTTVYRSGPSPINFA
```

gene

```

Query Match
Best local similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 gtgacgcccagcatt 20
Db 48632 gtgacgcccagcatt 48647

RESULT 27
AC018267/c
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
```


pieces.
AC018267
AC018267.1 GI:6552924
HTG: HTGS_PHASE2.
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 60573)
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10214125 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1.60573
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT
16964 a 14192 c 13649 g 15768 t
ORIGIN
Query Match 64.0%; Score 16; DB 65; Length 60573;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0y 5 gtcacgcgcagcatt 20
|||||
Db 27002 GTCATCGCAGCATT 26987
|||||
RESULT 28
LOCUS ATAC009895 82289 bp DNA 24-JAN-2001
DEFINITION Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,
complete sequence.
ACCESSION AC009895
VERSION AC009895.4 GI:12408719
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 82289)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.,
Ronnig,C.M., Koo,H., Fujii,C.Y., Ullrich,T.R., Barnstead,M.E.,
Bowman,C.L., White,O., Niernm,W.C. and Fraser,C.M.
TITLE Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 82289)
AUTHORS Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 82289)
Lin,X.
Direct Submission
Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Jan 24, 2001 this sequence version replaced gi:12280866.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org

BAC clone T21P5 is from Arabidopsis chromosome III and is near the
molecular marker ml172.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cds.dtu.dk/netgene/cdsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
location/Qualifiers
1.82289
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
/map="ml172"
/clone="T21P5"
1.24217
/note="overlap with BAC clone T12J13
(AC009327:65718..89934)."
complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/gene="T21P5.1"
complement(8..>3100)
/gene="T21P5.1"
/note="similar to hypothetical protein GB:CA838918
[Arabidopsis thaliana]"
complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
2979..3100))
/gene="T21P5.1"
/note="unknown protein"
/codon_start=1
/protein_id="AA01600.1"
/db_xref="GI:6017117"
/translation="MSQISGNNNIPLESEYVWLVNKADKRSKINDLPYERAYEN
YFPEKVEVYVYOLKMFQDNRKLVKMEIGELASRAQLYGHYMTSDAGLS
EYVEYFALITREFEKDGLFDINANKRQFLAPFLAMVLVIGREHVAQLVDQFKR
LIDCKRTROETPEKMKVVAQELIVRFLKSDTAPMNIPLRYSLVLDPNDACTPAS
RSLRLDALISYYCNKYSSEITLDSFRMLDLEPSSGLYQSGAKGONAPGV
ARINSOSMNDPLPAPKAVLTPSTLTHLAVLATICEELPBGILLILYLSASGIG
QISSPSARSATSVENILRDESHIKETPSQITPSGSSYIYPSDLVFTKPK
LFIIIDSSVFNKICGAEGPAAILLSPFTPLISAFSPRQPSGSLFTFTSP
VQAFCLSEISVKNMENDITFKAEKLLSSMNMWASTLATSDILHVSQILNDPFLR
RLILRFECRAVALIYPPVNNQNPDECPSPLESLPTPAVQSAVFQMANVFGAT
SKFTIPDDITMLSEF"
join(<3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..>6144)
/gene="T21P5.2"
<3627..>6144
/gene="T21P5.2"
/note="predicted by genscan"
join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)

```

/gene="T21P5.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AA01580.1"
/db_xref="GI:6017097"
/translation="MDNRSSSESRKHEIEKDTJASRKLEDTNKL1QDPEMALVAK
VRSOEELHSLQERIAACLEKMDLNLKYEGERKADIRVAIDKONAEVTSALNEL
ARRKGDLEENLKLAHDLKVEDEERYIFWTSILGILAEYGPVANAATASIGKHLH
DLOLWKTACANDRIKELSSIVENOPGTDFISKDNHDPNSKTOASGSDNDYOTN
EOLLPMENTRNPNHIMODTESIRFENNOIGGSGGIFEPQPKRENGCYPLSSVAGKE
MIOEREEKAENSMMDAYNGNEEFASHYRESGPCIDGFOITGDAIPGKVLGACFPVR
GTTICQPMQWVHIEDGTRQYIEGATHEPTITADVDKLTAVECTPMDODGROYKYBD
ESGTSFNESESVSKDVLITMOGEIVLEFANDONKIRCGNVENG1"
/misc_feature
7253..6289
/note="exon predicted by xgrail, quality marginal"
6384..6503
/note="exon predicted by xgrail, quality excellent"
6794..6839
/note="exon predicted by xgrail, quality excellent"
6928..7033
/note="exon predicted by xgrail, quality marginal"
7293..7411
/rpt_family="(TAAA)n"
complement(7488..7547)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
repeat_region
complement(7654..7784)
/rpt_family="(TAA)n"
complement(7819..7864)
/misc_feature
8054..8113
/note="exon predicted by xgrail, quality good_shadowexon"
8263..8380
/rpt_family="(TAAA)n"
complement(8592..8627)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
marginal_shadowexon"
<8652..>9527
/gene="T21P5.3"
<8652..>9527
/gene="T21P5.3"
/note="contains zinc finger motif, C3HC4 type (RING
finger)"
8652..9527
/gene="T21P5.3"
/note="unknown protein"
/codon_start=1
/protein_id="AA01602.1"
/db_xref="GI:6017119"
/translation="MGSTGNPNMGTTYSYRDCSQGVGVCPQMCYVIFPPPSFY
LDDEDSSSDSPFLIALIGLASAFITVSYTLISKYCHRRHNSSTSAANIRI
VCSDEYMGCTNNNGATNPNTIGGGGGLDESILKSTVYKRMGMGPVSSSCS
VCLSEFOENSLRLPKCNHAFVPCIDMTLKSNSCPLCRATIVSSAVEIVDLTNQ
OIVENNISTGDSVYVYVNLDENSRSRNEYNEGSTPLXGDAGKRRRRRAETGVVK
LRRCCIDRRYTCGP"
complement(10838..10896)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
11174..11254
/note="exon predicted by xgrail, quality good"
11690..11700
/note="exon predicted by xgrail, quality marginal"
join(11814..12256,12381..12578,12676..13085,13162..13332,
13418..>13761)
/gene="T21P5.4"
<11814..>13761
/gene="T21P5.4"
/note="similar to unknown protein GB:AA032238 [Arabidopsis
thaliana]"
join(11814..12256,12381..12578,12676..13085,13162..13332,
13418..13761)
/gene="T21P5.4"
/note="unknown protein"

```

```

misc_feature
7253..6289
/note="exon predicted by xgrail, quality
excellent_shadowexon"
14893..14935
/note="exon predicted by xgrail, quality marginal"
join(<15141..15586,15675..15872,16258..16667,16780..16950,
17036..>17427)
Query Match
Best Local Similarity 100.0%; Score 16; DB 13; Length 82289;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 aggtgcacgcgcagca 18
Db 30302 AGGTGCATCGCCAGCA 30317
|||||

```

```

RESULT 29
LOCUS HS272E8/c
DEFINITION Human DNA sequence from clone 272E8 on chromosome Xp22.13-22.31.
Contains a pseudogene similar to MDM2-Like P53-binding protein
gene. Contains STSs, GSSs and a CA repeat polymorphism, complete
sequence.
ACCESSION Z93929
VERSION Z93929.1 GI:3425887
KEYWORDS HTG; CA repeat polymorphism; MDM2-Like P53-binding protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 117968)
REFERENCE
AUTHORS Graham.D.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
COMMENT
Clone requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 272E8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 410B11 (286063) is at 117869 in this
sequence. The true right end of clone 390N22 (AL008711) is at 61633
in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/MGP/ChrX

```

272E8 is from the library RPII constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR.pcrPAC2>.

FEATURES

```

source
    location/Qualifiers
    1..117968 Homo sapiens
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /map="p22.13-22.31"
    /clone_11b="RPII-272E8"
    /clone_11b="RPII-1"

repeat_region
    /note="MER20 repeat: matches 2..218 of consensus"
    /complement(3645..3946)
    /note="MER33 repeat: matches 324..2 of consensus"
    /complement(3864..3994)
    /note="MER3 repeat: matches 131..1 of consensus"
    4194..4555
    /note="MIR1A1 repeat: matches 5..365 of consensus"
    5433..5531
    /note="MIR repeat: matches 87..188 of consensus"
    /complement(6029..6111)
    /note="MIR repeat: matches 188..103 of consensus"
    6205..6322
    /note="MIR repeat: matches 106..256 of consensus"
    7754

unsure
    /note="predicted CpG island"
    8261..9521
    /complement(9741..9813)
    /note="MIR2 repeat: matches 146..74 of consensus"
    /complement(11533..11655)
    /note="MIR2 repeat: matches 145..74 of consensus"
    /complement(18288..19413)
    /note="LIMD2 repeat: matches 1088..5 of consensus"
    /complement(19266..19525)
    /note="L1 repeat: matches 5390..5135 of consensus"
    /complement(19531..22106)
    /note="L1 repeat: matches 4931..2303 of consensus"
    /complement(22115..22419)
    /note="AluX repeat: matches 301..1 of consensus"
    /complement(22424..22641)
    /note="L1 repeat: matches 2314..2093 of consensus"
    /complement(23516..23641)
    /note="MER25 repeat: matches 2036..1908 of consensus"
    25175..25396
    /note="match: GSS D82600"
    27948..28235
    /note="AluY repeat: matches 1..301 of consensus"
    /complement(29865..29020)
    /note="THE1B repeat: matches 364..1 of consensus"
    /complement(29889..30189)
    /note="AluIO repeat: matches 302..1 of consensus"
    /complement(31338..31487)
    /note="MER5A repeat: matches 155..4 of consensus"
    33121..33328
    /note="MIR repeat: matches 63..262 of consensus"
    33424..33723
    /note="AluX repeat: matches 2..301 of consensus"
    35811..36234
    /note="L1 repeat: matches 1575..2001 of consensus"
    36723..39034
    /note="L1 repeat: matches 2121..4429 of consensus"
    37893..38188
    /note="AluX repeat: matches 1..302 of consensus"
    39040..39339
    /note="AluX repeat: matches 1..301 of consensus"
    39353..40296
    /note="L1 repeat: matches 4449..5390 of consensus"
    40148..41111
    /note="LIMD1 repeat: matches 1..968 of consensus"
    /complement(41972..42157)
    /note="MIR1 repeat: matches 541..354 of consensus"
    /complement(43606..43671)

repeat_region
    /note="MIR2 repeat: matches 134..71 of consensus"
    45479..45585
    /note="MER45 repeat: matches 26..133 of consensus"
    48242..48312
    /note="MER45 repeat: matches 98..176 of consensus"
    49604..49826
    /note="AluIb repeat: matches 82..301 of consensus"
    49936..50472
    /note="L1P1A3 repeat: matches 41..573 of consensus"
    /complement(50932..51134)
    /note="MER20 repeat: matches 218..18 of consensus"
    51378..51596
    /note="MER20 repeat: matches 1..215 of consensus"
    /complement(52293..52552)
    /note="MIR1C repeat: matches 466..186 of consensus"
    /complement(52555..52908)
    /note="THE1B repeat: matches 362..1 of consensus"
    /complement(52914..53071)
    /note="MIR1C repeat: matches 182..8 of consensus"
    53212..53621
    /note="MIR repeat: matches 1..403 of consensus"
    55202..55295
    /note="MIR2 repeat: matches 1..94 of consensus"
    55296..55504
    /note="MER3 repeat: matches 1..209 of consensus"
    /complement(55783..56684)
    /note="L1P1A10 repeat: matches 911..1 of consensus"
    /complement(56534..57144)
    /note="L1 repeat: matches 5390..4785 of consensus"
    /complement(57148..59003)
    /note="L1 repeat: matches 4720..2850 of consensus"
    59735..59757
    61714..61945
    /note="match: SMS L24834"
    /complement(63228..63358)
    /note="MIR repeat: matches 247..100 of consensus"
    /complement(64335..64455)
    /note="LIME3 repeat: matches 909..790 of consensus"
    /complement(65455..65744)
    /note="AluIO repeat: matches 289..2 of consensus"
    66343..66557
    /note="MIR repeat: matches 41..261 of consensus"
    66912..67656
    /note="LIMC3 repeat: matches 1660..2409 of consensus"
    /complement(68490..68794)
    /note="AluY repeat: matches 301..1 of consensus"
    /complement(68873..69190)
    /note="AluX repeat: matches 298..1 of consensus"
    69217..69532
    /note="MER33 repeat: matches 1..324 of consensus"
    /complement(69638..69703)
    /note="MIR repeat: matches 129..64 of consensus"
    71210..71340
    /note="MIR repeat: matches 83..206 of consensus"
    72426..72792
    /note="match: GSS A005396"
    /complement(72923..72973)
    /note="MIR repeat: matches 139..89 of consensus"
    /complement(75007..75196)
    /note="MER5A repeat: matches 189..1 of consensus"
    /complement(75386..75425)
    /note="LIME3 repeat: matches 406..367 of consensus"
    75508..75800
    /note="AluIO repeat: matches 1..288 of consensus"
    /complement(76744..76840)
    /note="MIR1 repeat: matches 182..91 of consensus"
    /complement(78342..78521)
    /note="MIR repeat: matches 216..35 of consensus"
    /complement(78368..79020)
    /note="MIR2 repeat: matches 144..112 of consensus"
    /complement(79137..79186)
    /note="MIR2 repeat: matches 134..85 of consensus"
    /complement(79151..79393)

```

```

repeatL_region /note="MIR repeat: matches 242. .3 of consensus"
                  complement(79492. .79666)
                  /note="MERSB repeat: matches 178. .5 of consensus"
repeatL_region /note="MERSB repeat: matches 178. .5 of consensus"
                  complement(79777. .79863)
repeatL_region /note="MERSA repeat: matches 109. .27 of consensus"
                  complement(79894. .79970)
repeatL_region /note="MERSA repeat: matches 11. .91 of consensus"
                  complement(80377. .80431)
repeatL_region /note="MIR2 repeat: matches 141. .87 of consensus"
                  complement(80389. .80462)
repeatL_region /note="MIR repeat: matches 252. .177 of consensus"
                  complement(80455. .80578)
repeatL_region /note="MIR repeat: matches 147. .29 of consensus"

Query Match      64.08; Score 16; DB 92; Length 117968;
Best Local Similarity 100.08; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atcgccagcattcga 24
    |||||
Db 97872 ATCGCCAGCATTTCGA 97857

RESULT 30
AP002817/c AP002817 140825 bp DNA PLN 12-AUG-2000
LOCUS      Oryza sativa genomic DNA, chromosome 1, PAC clone:P0699D11.
DEFINITION AP002817 BA000010
ACCESSION  AP002817.1 GI:9558510
VERSION     AP002817.1 GI:9558510
KEYWORDS
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0699D11.
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
            Oryza.
REFERENCE   1 (bases 1 to 140825)
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0699D11
            Published Only in Database (2000) In press
JOURNAL     2 (bases 1 to 140825)
REFERENCE   Sasaki,T., Matsumoto,T. and Yamamoto,K.
AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE      Direct Submission
JOURNAL     Submitted (26-JUL-2000) Takuji Sasaki, National Institute of
            Agricultural Resources, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@agr.affrc.go.jp, URL:http://rqp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP(PfR,SWISSPROT,
            GENEPIR, PDB) from MAF DNAbank and the cDNA sequence database at
            RGP. Protein similarities of the coding regions were searched
            against NRP with BLASTP2.0. ESTs represent the identified cDNA
            sequences using BLASTN2.0 with the corresponding DDBJ accession no.
            and RCP clone ID.
            This sequence of P0699D11 clone has an overlap with P0462H08 clone,
            DDBJ:AF002525 at the 5' end. This sequence of this clone starts at
            the position 114,739 of P0462H08. This sequence of P0699D11 clone
            has an overlap with P0469E09 clone, DDBJ:AP001366 at the 3' end.
            The sequence of this clone ends at the position 59,396 of P0469E09.
            Detailed information on assemble quality together with annotation
            of this entry at
            http://rqp.dna.affrc.go.jp/Genomeseq.html.
FEATURES
    source
        1. 140825
            /organism="Oryza sativa"
            /cultivar="Nipponbare"
            /db_xref="taxon:4530"
            /chromosome="1"
            /clone="P0699D11"

CDS
    join(527. .679, 925. 1114, 2076. .2467, 3067. .3555)
    /note="EST C22619(S11214) corresponds to a region of the
    Predicted gene.
    Similar to Arabidopsis thaliana chromosome 2, BAC clone
    t17D12; putative protein kinase (AC006587)"
    /protein_id="BAB03429.1"
    /db_xref="GI:9558511"
    /translation="MWGIRRNNGHEATITFLRELADNTNNESTECILRGQFSSVYK
    APLNDROYVAVKOLDINGLQGNREFLVEVLMSTLHHPNLVLFQYCVADGQRLIYE
    YPLGSLERLDLHLPQGEPLDTRRMKIAADAAAGLELHDEALPAVYIDIPSN
    LLEGYNKATSDPGLKAGVGDKTVTTRGTGTCGACAPLEYLSTGKLTISDIYSPG
    VVFLDLITGRALDSDNRPPEQDLVAMARPLFDORRKPAAADPSLHHPKRGIFOA
    LAIAMLCOEKAKNRPSTREVEVALSYLAOFHESONTAAHHTLPGSPVLDVNOIY
    ODTSLPSOHGVMPPLAGDHDHVVQVCKENCSSSHIRPRGVTTPMGADREARALAEANV
    WYEARROKTSKMR"
    complement(8769. .9326)
    /note="ESTs AU093260(C52190), AU062991(C52190) correspond
    to a region of the predicted gene.
    hypothetical protein"
    /codon_start=1
    /protein_id="BAB03430.1"
    /db_xref="GI:9558512"
    /translation="WDFLEPTLEICRAFTIRGISDKYALSKPALVDASMSI
    AVDQSCVMDICDFEGNMGIFLDLVGSSDPHISDASCLVLLKAPREAVTGLTN
    LPLKLSVLDLKHGTCRLTRILYCLAFSCQYLAQGVISISLSALMRVALVSARK
    GSHDCLADAASYLCAELORLPRCG"
    complement(join(15587. .15727, 15868. .15915))
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB03431.1"
    /db_xref="GI:9558513"
    /translation="MNNFHYVEALIGRCKHSTVYKGRKKSLIEFAVAVSDKQSKVL
    NEVSAAPSSPCSFISSW"
    complement(join(18733. .18801, 19343. .19489, 19843. .19905,
    20016. .20087, 20611. .20691, 20769. .20819, 20976. .21072,
    21611. .21753))
    /note="ESTs C72327(E1437), AU078752(E1437) correspond to a
    region of the predicted gene.
    Similar to Archaeoglobus fulgidus competence-damage
    protein (AE000949)"
    /codon_start=1
    /protein_id="BAB03432.1"
    /db_xref="GI:9558514"
    /translation="MISDGRERAGRTKKNISSVSSNGTSTVEQITSKASIIYVG
    DEILLEGVEKLGAGLCKLHAIQWRVSHAVVSNEDSYAEVEECKSTDDWFLVG
    GLGLHSDISLACVAKAFGVRLAPDEFEYLSQLIGDNYTCDRNEMALLPEGITEL
    HHKMLPLPLIKCKNVILATNDVDELETWEGCLLDQESGLVNAKSPVSKHICTSLD
    VKTAPVYAKLCIDPSDVYIGMT"
    join(30551. .30609, 30919. .31139, 31223. .31578)
    /note="Similar to Arabidopsis thaliana chromosome 4, BAC
    clone F10M23; hypothetical protein (AT035440)"
    /codon_start=1
    /protein_id="BAB03433.1"
    /db_xref="GI:9558515"
    /translation="MLGKEKRRREGGENDIYAKKSMGPDPLPHIRGSISIKVP
    QIVYIGTKNVLEPQNRKTAGSLNCTKPTRRSSVLTPLVLYVAESMNFRRHKKRGH
    HGDLRQAGAAERDEEVYTDGLDHRQPHVKIPSEAAVYARRRMRMTQSSVGED
    GDAAVSPHKVAVARRAAHSSVLEAGRMKGLCDLHNRNVLCTGGLD"
    complement(join(31912. .32455, 32698. .33684, 34091. .34341))
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAB03434.1"
    /db_xref="GI:9558516"
    /translation="MRKKRTNLLIVGEGATFEFFEEERCRRRAGVSGSEVERRGDSKP
    SRGHRPATARAPLRRQKARREDELTAASSGWTPESGWRPHHPLPSPLLSLSH
    SSLIFSGRLGEHOCGEAASGSGVSAAGSGAATAAAGAAAGCGCCGCAAPLSP
    RHPALSSAAGPARGDEDAGLVGSGEGGTGAGARARRRQIOGERRHMKRKPVS
    VCCCFEYVAELVYLAIVCVPAAIHLAVRGGRHRRVCAKQKEMGELLTLDVTPRS
    LAAAAKARKVVEFPATPMABHIGAEVARRKSSLOALHIVYAKLLSSAGDARPA
    DLRLVLYLACPLSPPTPLPRIPRHRSLSLPRKLASLHIVYAKLLSSAGDARPA
    TEGNKTIKERENSIITAOETCRSRAPATMPPTVARRAPNPDSRWGGAAPASAR
    EBGGGGGGDAALPACRLTAAYSLAASPKVTPPKPAVRPLTNHPNIRORAPQRCR

```

```

CDS
SSPSLSSSSLSGSSGKPARCGTVARERLFWRRVYQLATSIHAFAAHVIVVVL
VKLPAREACPRSTAPASAROGPPHRRR"
/join(35382..35472,36089..36172,36426..36530,37078..37337,
38995..39114)
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB03435.1"
/db_xref="GI:9558519"
/translation="MQLDASDFHGOELTLKLRDGLKLSLESSGKSYDLVFPAAQOP
DAVFLPSCGEAAVGIARVSLVRYPPDELEKPGLSLTPSSKRSAGSSKTKRSR
FTSGSKNRSSGSAOGLGOOSAEPAHKMKRDESLGHSNVSAGSAGSOVRGDS
GTSSEVPQTPEVKKKKKKKKLDPMKRRTPYRHHKTAQDLDRTAVDVSSLYKSPKPY
Q"
/join(40731..40753,43157..43538,45516..45821,47347..47893,
48657..48934,48983..50395)
/feature="Similar to Arabidopsis thaliana chromosome II BAC
clone T27A16: hypothetical protein (AC005496)"
/codon_start=1
/protein_id="BAB03436.1"
/db_xref="GI:9558518"
/translation="MYSLRPKIRPPFALAAADSLPMPPPPPGRPLPSPWRGRDR
RLLPYASRSGPREGSQRGRENEEEREREGEGEGEELKGRGKEERE
EIRLQAGMRMRACGTFADFLSRLNNKSLVWLIADETLKOLAGNRDPIGIG
HGAARDLPIRDSIYAKLAGHSIQHLSGSMHLGPCSCAGLMRRRSKASASAL
LISGCMNPOTPCPNRSORPATTLATISLILSPPEFLSYELONHYSCIFISI
ESLNLGRRSPSPPTSTVPPMTATVRSIYIIPYDGVREMRKLAVYASISS
SLNPMEDOTVRCIGSSALSIGITPIHRAKPIAFGCFNPFPLKYSTALAH
AYAHSSSINKRAHPTRIKLIDILYKQAPNLCTSPCRNRSNHNHSAQCTNRLPAQ
YRVLDSEAHPPKQREERARETARPHDRSLSPNPPAPATERRPPPSLP
DLRAHAYLLPRGHRHARVASALLASLRLPRDAALVRRVHVSLLAAARLGG
RGGGLAAOHLHLVRAGHAADPHASASIVAOVSCGSAVASARVDETAASDVVSM
NWKIDYVSGDLARARELEDVWGRNVSMPVIGAYAKOMQPEATVFRMOVGE
IIPDGYALLSVLSACDGLGVYDIGEVHVRVLRGICQELPLMNLITIMYVKGSKK
ALEVPEGMESKIVTWMTMAGFALHGLGLEAELPRMEKVENSDNITDAVLSVC
SHVGLDGLRWYFKTWTWQYKTKPRVHNGCMIDLLGRAGCLMEARGLIDMPFKNA
AIGCALIAARTHGTDELGEOLHLEIEPHNSGVYIILSNYARDEMDVARELRI
SMRDGLRNYPGASIDVGMVHEFTSRDGSHTLAKIREVLCAINSIKSVCHILL
PESLHYEEB"
complement(join(52441..53547,53887..54014,54128..54165,
54272..54487,54709..54870,55838..56016,56150..56256,
56375..56539,56658..56741,56821..57040))
/feature="Similar to Arabidopsis thaliana chromosome I BAC
clone T5A14: kinesin (AC005223)"
/codon_start=1
/protein_id="BAB03437.1"
/db_xref="GI:9558519"
/translation="MDVQPARTMRNLPTLTLMSGFNKHLTPSMIESVHIIDGLSPT
KPMKVAVKEDENISDNTSEAKVOKIODELVLNAOLKQITLQREALNNYLDLG
NIRVFCIRPEHHESEYSRNLTFLDESNNVLEAVETKRQYKEDVFPQDSQGVF
SEVEPVKSAIDGYNVICIFAYGOTSGKTYTMGKPTNLGIVTIGTQTLFNQASGSE
RFLFTSMELTYMGNIRDLAPRSKINGIKNVSLSIKSPDGCITEIDLVATVNSF
QEVRLTEWGTRLRSTASTMANSTSSRSKHLIRISLISLWATERKATSKLWMLDGG
SERLVKATGKGRKEKGAINTLSLADGVIDALQTKPHVPRNSKRLTYLIDLSGC
ESRTLMVHISPDGDCETICTLGFATPRSLRLESEPEPKAKRETLIDLGQV
NDLEHEDCIRRRKIKNLESEMEHITGPOPTIYNLPMKSHLSSELTKDVSANRNSK
RRSSRLPRPMKPTASSOHRGINNRTPIINLKPPVRRRRSPVSVASVAVPNA
APMOSSESSSECSMLTSDMMTPSIRDTGCSODASYTEIKQVTFSEHKSMDQYTC
YTDPLAESKDIOIKTEKGIVIDNNLHOQVTEKSTFSKAWLDPGTAEHLVHS
SIPSPITMACTKEDSQVDEVMGLTLOSTDYVEDIKOSTDQOFAKELCTPFEKE
SSNNEVGHKNEHVPYHGRPPRSLOELENCTLEKPMNDSKSHRSHDKTKTGNVLSY
TGE"
/join(57603..57645,57742..57843,58195..58344,59425..59479,
59537..59589,61304..61785)
/feature="hypothetical protein"
/codon_start=1
/protein_id="BAB03438.1"
/db_xref="GI:9558520"
/translation="MKQDKNNRCLTFLGSGSAPYIHOVITIGKEGCPDTPSPRDSR
SEYAKIMIDYPMHMTTPNGMNCLANENHRSRAVIRKSGSPMCHAEJLSEPTAAS
LIALSLAGKAFELSKRRSGDVHMLAISCVEAAAPLPLQPPILPPDNDRAAGG
CLASPASARRLPPPLPHOROPMDRAAPAAAGRLSPNARRRELKRPPLPISTSP
AIGRCMRPPLADATGELVELVSSKCRSCVETIPSTVELLPMVVGELRFLFSYVL
LPSPSCPAMSLAPSSRA"

```

CDS

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9: hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

Query Match 64.0%; Score 16; DB 13; Length 140825;
 Best local Similarity 100.0%; Pred. No. 36;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcccagcatcttg 23
 |||||

Db 89622 CATGCCACATTTTCG 89607

RESULT 31
 AP001366/c DNA PLN 12-AUG-2000

LOCUS AP001366 146081 bp genomic DNA, chromosome 1, PAC clone: P0469E09.

DEFINITION AP001366 BA000010

ACCESSION AP001366.1 GI:7228436

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

CDS

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9: hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

LTR

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9: hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

CDS

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9: hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

predicted gene.
Similar to maize transposon MudR mudra protein isolog
(AC003981)"
/codon_start=1
/protein_id="BAA92398.1"
/db_xref="GI:7228438"
/translation="MTSKGIDRPAERFPOSITYSWLLGFRIDQEVYTMSSVVSRA
EGYFELMFMNSTAMRRYVYMAEERFPIVSVQEDLINVMOTEDDEGPINAD
VVGFSMONEOPRECAMGADGERGIIIVDEREDSDNEPADDASSDEGDVMA
KTDVNEDESGVILLEGDHPWEYKENEVIDGARYAKDEMEKAEVAKMAVSLQEEFV
ASTNYVEVCKMEDCEPMVHAYKGMNDWKSIVTIEKKYLOGEYKHNITSEF
VASEYSSVGNIGPEEYSIIIRHTEENKRYISVAKCAKAKOITIEBMYGFEASYN
LPRLIATVQNNNTYVDLHFTSDVDRTKSVIQRAPSLCACINAYHCPVCLIC
TEMTKRGQGLTIAIGDGNNOVYPAHAPESNTSSWTFLERLITAVRMKPNVC
LIHRHAGMLKELIDNGDEKLPKAPKDPVRSWCMRHGANGYKQFNKHLMLF
KRLCAKNQKEAFENLMDKLDLTKQDEGRSROVEGDEPIPLGALHDDPPMRK
SGSSIIRNFTQWIEPEKESLLEDTDSRGYIMTNI AEYNNVMGVRVLPVAVI
EFLHTQAYFRDYKKIGPSMADNNIVFGEQVRLPLHGAPANTLPYSSVGPGRAM
LLOFTADLHPDADNNSVRSLEAYILMFGWMTSTHGHANDPRLVHYARSTYDO
PODYPOWSGSAVIAATYRALCEACTTDCAGITACCPMLQMAERFATGRVDS
APYVGSACQPEGPTMGTYWCRGRRTYAVOVRGYPDVEFEEDLQPSDVIWERY
TEEVAARAPLGSLCTRDQAWLTLIPWFELFEPHCFORVMRQGFROYEPGV
QPTVLPADHSLTRRGOLAGALMARVOOYVDWVLATEEYINELFPHTEENYDYLEW
YLPRTRAVITTPAPRPHAAVTDAYPIHRDXYFVGDAARDISADITAVOYLR
GLHLDVEOARTFDRMOEKRAVMRVFSCSAVITVPAGPDLVCLRAPLASOCCD
HFRFTYRPMCNFTIEWDMENPDGHRAYRSTRSYLRKDEHCRITAAALEH
VNLGLPTMRREPCRCGRDCHYIRSVRHTTGRGRCVFNIVDDDFVTHAMEEYR
RVELASIDMRNRPGLFRKWSBRPCGRDMOYATSPAEOTFRRYFVCPNIDFVY
CLTIYKIRYDLHKITTYTE"

LTR
/note="3' LTR"
join(23418..23748,25558..25589,25715..26480,26838..26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92399.1"
/db_xref="GI:7228439"
/translation="MGANCCITAKERPQCVTPLEVSAPFNVRHSPSMFRMDNTHI
EDIMEMALFSNHSSTSIPEETKSGSIADPDGFSNGSGSPDMFNKILCHKSDKRSS
KIRSLDRGSHSSSSSSPEAKISRSKIDYVSAVSDSKMSISVSTPAISRADPSSS
RGSILPTDASMKARSPGYOLROYSDSKTISLRKLNKNGASPEGSSSLVCSN
DLASVSGHSGSSDSMSKRTSEMAVSORERMSVDSLIGSVSKMTRMANSNNPTH
SPDQVCKLCLLKERSTWNAOELAAVAAVLGLGHYHADCLDLSLTAEDAKPPCPV
CTHEKCTGAKLFGLESRTKNKIPKNIYDVNLDGSKHOKKIKPLEEELYSQARASM
RELKSSRIGANDMTLNDLIRIEGS"
complement(join(27953..28624,29565..29918,30369..30666,
31026..31191))
/note="ESTs C96615(C10106),C26336(C12127),D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (AL021889)"
/codon_start=1
/protein_id="BAA92400.1"
/db_xref="GI:7228440"
/translation="MESLRDMVLPGRGFHPKTELISHLKKTKTGOKLEFIIPEY
DIYKHEPMDLAKCDVPTQNMKHFPAARDKIPNSRSNRATVAGYWMSTGDRKIK
MGOTCTKTTLVFHGRPTGRKRTIMHEYIDRECOQCPMDKADVLCRTKRN
DWLPGNGENLNSDPHEPEYDAPSVISTEQLMDPAEPVYGAAPVTADEPGCVTS
AITANLPSPSDILNDLWNLNLFDPFDEQSLASDLSPDEQNVSSNVALAPVE
QDYSPNENNVDDTEYLEDPEYNIILPGCTDMDMONPLDPPLOYADVNSGJKE
ELMSPOANAPESQSEADANGILIRYRSKMTPTSVOPKGTQAKMRGINKMASS
SESINOTIKRENSGRIVHOKNOAHVASTKREDAKPTSELSSNGFLRGIRNAPAG
CSADRNMILVAFAGVAVVAHICORGLSORDOHT"

CDS
complement(36747..38998)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92401.1"
/db_xref="GI:7228441"
/translation="MMGHPIYRADTARPPNNMAVPRPEYRHVDRHCTPVPVIVLCRPD
DRGPCASLMAHMLGELVHGCCCTDDEPHGPGSTI"

LTR
/note="3' LTR"
complement(39315..39726)
complement(join(40209..40396,40917..43545))
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.

Similar to maize transposon MudR mudra protein isolog.
(AC003981)"
/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSSKVTFOIVHGEINIRFGPDGVDLSDFVMTSKGIDRPAERTFO
SIVSWLLRGRIDQEVYTMSSVVSRAEGYFELMFMNSTAMRRYVYMAEERFPIVSVQEDLINVMOTEDDEGPINAD
VVGFSMONEOPRECAMGADGERGIIIVDEREDSDNEPADDASSDEGDVMA
KTDVNEDESGVILLEGDHPWEYKENEVIDGARYAKDEMEKAEVAKMAVSLQEEFV
ASTNYVEVCKMEDCEPMVHAYKGMNDWKSIVTIEKKYLOGEYKHNITSEF
VASEYSSVGNIGPEEYSIIIRHTEENKRYISVAKCAKAKOITIEBMYGFEASYN
LPRLIATVQNNNTYVDLHFTSDVDRTKSVIQRAPSLCACINAYHCPVCLIC
TEMTKRGQGLTIAIGDGNNOVYPAHAPESNTSSWTFLERLITAVRMKPNVC
LIHRHAGMLKELIDNGDEKLPKAPKDPVRSWCMRHGANGYKQFNKHLMLF
KRLCAKNQKEAFENLMDKLDLTKQDEGRSROVEGDEPIPLGALHDDPPMRK
SGSSIIRNFTQWIEPEKESLLEDTDSRGYIMTNI AEYNNVMGVRVLPVAVI
EFLHTQAYFRDYKKIGPSMADNNIVFGEQVRLPLHGAPANTLPYSSVGPGRAM
LLOFTADLHPDADNNSVRSLEAYILMFGWMTSTHGHANDPRLVHYARSTYDO
PODYPOWSGSAVIAATYRALCEACTTDCAGITACCPMLQMAERFATGRVDS
APYVGSACQPEGPTMGTYWCRGRRTYAVOVRGYPDVEFEEDLQPSDVIWERY
TEEVAARAPLGSLCTRDQAWLTLIPWFELFEPHCFORVMRQGFROYEPGV
QPTVLPADHSLTRRGOLAGALMARVOOYVDWVLATEEYINELFPHTEENYDYLEW
YLPRTRAVITTPAPRPHAAVTDAYPIHRDXYFVGDAARDISADITAVOYLR
GLHLDVEOARTFDRMOEKRAVMRVFSCSAVITVPAGPDLVCLRAPLASOCCD
HFRFTYRPMCNFTIEWDMENPDGHRAYRSTRSYLRKDEHCRITAAALEH
VNLGLPTMRREPCRCGRDCHYIRSVRHTTGRGRCVFNIVDDDFVTHAMEEYR
RVELASIDMRNRPGLFRKWSBRPCGRDMOYATSPAEOTFRRYFVCPNIDFVY
CLTIYKIRYDLHKITTYTE"

LTR
/note="5' LTR"
join(44572..44723,44846..44938,46366..46378)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPGRVSRPKPMAAAAASAPROPBGHMLVLTGTVGRVSO
RILAQWRVSGTCTSPAKTELEMLGMDASIFDRTSSCWM"
complement(join(48119..48257,48429..48703,49975..50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALVNDKSDSDVNAAPCARAPVILGAPMGVYLOLARLUGEC
ERDRARMAIISFCAISLBSKTRITYVILCYLRNVPPLVTESKRTKCRIT
ASLARSDELVCICRAKLASSALDRVFKLITSDPVLWLKRCICVINPVHSRTH
GSSMYSTLPDAVPYRSTRLAGVICTIGC"
complement(join(51918..52637,52664..53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUP8H12R sequence, C2-HC type zinc finger protein
Cε-MYTL (AC002986)"
/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="MASCINGELHVRVYHRLVKASDESIRPHVLAIVNSDILPRTI
OVSFICIKPSTGDFHDVVAFAVAGPSLLNHPPLAGIVSNPGSLPEICHNO
GAEIVGADVAIASIDYGTGASVYKTLIPYAGDVLAISVYGFACGPGFVANGTNR
VYVDSASLSMYSANSELARSGTILAAGRPHHDSVRRSPSTGASLDRAFTPLG
AROADILKREOKATRVQAVSATLWMLAAVGSRRARCRMMWMDVGRRLTLSS
PELRAMRSYVGNVTTFVAVATMEELORKIABVAAVMAADAIAPAVGEHFDLVW
VEHRAGQRYIDTASVGLGSPAVSVTASFVDTDFGHGHAAMALPTSSSARLCT
GFVOIARPGGDSWIASALLMPRLAALESDBRLGRIFPVAYEVLGRPAGSSA
KRAGVITKTI"
complement(57570..58009)
/note="3' LTR"
complement(join(58018..58891,58980..62098))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F9013 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"
/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="MRVNLQAGGLMTAIDPEYAEFREDMAALSAILOAVREMLRGLA
KHDTAKAAMDAIKTRMGVDRVREAKGQSRROGSRFRERFEFARLTAIVAD
IRDMGVMEDEHVKRLILRVPRKKYKVALISLELDLVKTMALIELYGRSLTSDYS
DEGSDGKILYLTQEOARVYKORQEGSGKGGKAGATGQOHRKPGSGKGA
ATGANSRDLISRVKCFMCDERGHYARCCRRPRORRGRALYQAADEEPTLLAHVYG
VSLGPAITLGRTPGGQVYHTEKKVILIDHEDGGEETRWPLDTIGATNMTGVSAT
AELDTGVGVTKFGDGSVITQGRGVVFCRKNDRSLDTVYIIPKLRNITISVGL

Query Match 64.0%; Score 16; DB 13; Length 146081;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 catgccagcattcg 23
 |||||||||||||||
 Db 8193 CATGCCAGCATTTCG 8178

RESULT 32
 AC034114/c
 LOCUS AC034114.4 154306 bp DNA HTG 17-FEB-2001
 DEFINITION Homo sapiens clone RP11-775E10, WORKING DRAFT SEQUENCE, 8 unordered
 pieces.
 AC034114 GI:12958060
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 154306)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens, clone RP11-775E10
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 154306)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Bonquist, A., Bouckhalter, B., Brown, A., Buckett, G.,
 Campilayo, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gargana, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, D.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McKean, P., McGurk, A., McKernan, K., McPheters, R.,
 Merdum, T., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (04-APR-2000) Whitehead Institute/MIT Center for Genome
 Research 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 17, 2001 this sequence version replaced gi:8781909.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4353
 Center clone name: 775_E_10
 ----- Summary Statistics
 Sequencing vector: M13; W77815; 41% of reads
 Sequencing vector: Plasmid; n/a; 59% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151316 bases at least Q40
 Consensus quality: 152394 bases at least Q30
 Consensus quality: 153049 bases at least Q20
 Insert size: 151000; agarose-fp
 Insert size: 153606; sum-of-ctrls
 Quality coverage: 10.6 in Q20 bases; agarose-fp

Quality coverage: 10.4 in Q20.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 41084: contig of 41084 bp in length
 * 41085 41184: gap of 100 bp
 * 41185 49589: contig of 8405 bp in length
 * 49590 49689: gap of 100 bp
 * 49690 58448: contig of 8759 bp in length
 * 58449 58548: gap of 100 bp
 * 58549 70162: contig of 11614 bp in length
 * 70163 70262: gap of 100 bp
 * 70263 85680: contig of 15418 bp in length
 * 85681 85780: gap of 100 bp
 * 85781 106804: contig of 21024 bp in length
 * 106805 106904: gap of 100 bp
 * 106905 139784: contig of 32880 bp in length
 * 139785 139884: gap of 100 bp
 * 139885 154306: contig of 14422 bp in length.

FEATURES
 source
 1. 154306
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-775E10"
 /clone_1b="RP11 Human Male BAC"
 1. 41084
 /note="assembly_fragment
 clone_end:sp6
 vector_side:left"
 41185..49589
 /note="assembly_fragment"
 49690..58448
 /note="assembly_fragment"
 58549..70162
 /note="assembly_fragment"
 70263..85680
 /note="assembly_fragment"
 85781..106804
 /note="assembly_fragment"
 106905..139784
 /note="assembly_fragment"
 139885..154306
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"
 BASE COUNT 43533 a 31887 c 31913 g 46271 t 702 others
 ORIGIN

Query Match 64.0%; Score 16; DB 71; Length 154306;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

6 tgcacgcagcattt 21
 |||||||||||||||
 Db 78184 TGCACGCAGCATTTCG 78169

RESULT 33
 AC002457/c
 LOCUS AC002457 177380 bp DNA PRT 21-DEC-1999
 DEFINITION Homo sapiens BAC clone CTR-60P12 from 7q21, complete sequence.
 AC002457
 AC002457.1 GI:3947433
 HTG:
 KEYWORDS human.
 SOURCE Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

repeat_region      21136..21171
                    /rpt_family="(CA)n"
repeat_region      21478..21940
                    /rpt_family="L1"
repeat_region      22310..23003
                    /rpt_family="L1"
repeat_region      23135..23653
                    /rpt_family="MER21_g"
repeat_region      23658..24140
                    /rpt_family="L1"
repeat_region      24263..24304
                    /rpt_family="AT_rich"
repeat_region      24419..24447
                    /rpt_family="AT_rich"
repeat_region      24618..24865
                    /rpt_family="Alu"
repeat_region      25627..25650
                    /rpt_family="(TTAA)n"
repeat_region      25906..26152
                    /rpt_family="L2"
repeat_region      26153..26536
                    /rpt_family="AT_rich"
repeat_region      26654..27556
                    /rpt_family="Retroviral"
repeat_region      28842..28903
                    /rpt_family="(CATTA)n"
repeat_region      29888..30080
                    /rpt_family="MER1_type"
repeat_region      30382..30669
                    /rpt_family="Alu"
repeat_region      32790..32960
                    /rpt_family="MIR"
repeat_region      33062..33192
                    /rpt_family="L1"
repeat_region      33199..33998
                    /rpt_family="L1"
repeat_region      34000..34042
                    /rpt_family="(TAAA)n"
repeat_region      34043..34316
                    /rpt_family="Alu"
repeat_region      34414..34602
                    /rpt_family="L1"
repeat_region      35064..35263
                    /rpt_family="L1"
repeat_region      36255..36571
                    /rpt_family="Alu"

Query Match      64.0%  Score 16;  DB 85;  Length 177380;
Best Local Similarity 100.0%;  Pred. No. 35;
Matches 16;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      3  aggtgcacgcagca 18
        |||
Db 48179 AGCTGCATCCGACGA 48164

RESULT 34
AC084136/c 180700 bp  DNA  HTG  14-FEB-2001
LOCUS      Homo sapiens clone RP13-506N4, WORKING DRAFT SEQUENCE, 32 unordered
DEFINITION
pieces:
AC084136
VERSION    AC084136.2  GI:12830241
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 180700)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
            Boucknight,B., Brown,A., Burkett,G., Campiano,A., Castle,A.,
            Chokipali,V., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            Fitzhugh,W., Gage,D., Galagan,J., Gaidyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Lacroque,K.,
            Lamares,R., Landers,T., Lechoczky,J., Levine,R., Lien,C., Liu,G.,
            Macdonald,P., Margulis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheters,R., Meldrim,J., Meneus,L., Minova,T., Menga,Y.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sounguez,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.

            Direct Submission
            Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Feb 14, 2001 this sequence version replaced gi:10800269.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu

            Project Information
            Project name: 111406
            Center project name: 506_N_4
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 167122 bases at least Q40
            Consensus quality: 176531 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 176000; sum-of-contigs
            Quality coverage: 3.7 in Q20 bases; agarose-fp
            Quality coverage: 3.8 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1      951: contig of 951 bp in length
*      952 1051: gap of 100 bp
*      1052 2064: contig of 1013 bp in length
*      2065 2164: gap of 100 bp
*      2165 3416: contig of 1252 bp in length
*      3417 3516: gap of 100 bp
*      3517 4816: contig of 1300 bp in length
*      4817 4916: gap of 100 bp
*      4917 6297: contig of 1381 bp in length
*      6298 6397: gap of 100 bp
*      6398 9144: contig of 2747 bp in length
*      9145 9244: gap of 100 bp
*      9245 11578: contig of 2334 bp in length
*      11579 11678: gap of 100 bp
*      11679 30905: contig of 19227 bp in length
*      30906 31005: gap of 100 bp
*      31006 32606: contig of 1601 bp in length
*      32607 32706: gap of 100 bp

```

```

* 32707 34256: contig of 1550 bp in length
* 34257 34356: gap of 100 bp
* 34357 37618: contig of 3262 bp in length
* 37619 37718: gap of 100 bp
* 37719 39131: contig of 1413 bp in length
* 39132 39231: gap of 100 bp
* 39232 41291: contig of 2060 bp in length
* 41292 41391: gap of 100 bp
* 41392 43172: contig of 1781 bp in length
* 43173 43272: gap of 100 bp
* 43273 45886: contig of 2614 bp in length
* 45887 45986: gap of 100 bp
* 45987 49186: contig of 3200 bp in length
* 49187 49286: gap of 100 bp
* 49287 52781: contig of 3495 bp in length
* 52782 52881: gap of 100 bp
* 52882 57288: contig of 4407 bp in length
* 57289 57388: gap of 100 bp
* 57389 62317: contig of 4929 bp in length
* 62318 62417: gap of 100 bp
* 62418 65388: contig of 2971 bp in length
* 65389 65488: gap of 100 bp
* 65489 70061: contig of 4573 bp in length
* 70062 70161: gap of 100 bp
* 70162 75771: contig of 5610 bp in length
* 75772 75871: gap of 100 bp
* 75872 83423: contig of 7552 bp in length
* 83424 83523: gap of 100 bp
* 83524 89716: contig of 6193 bp in length
* 89717 89816: gap of 100 bp
* 89817 97026: contig of 7210 bp in length
* 97027 97126: gap of 100 bp
* 97127 104654: contig of 7528 bp in length
* 104655 104754: gap of 100 bp
* 104755 114340: contig of 9586 bp in length
* 114341 114440: gap of 100 bp
* 114441 126202: contig of 11762 bp in length
* 126203 126302: gap of 100 bp
* 126303 142208: contig of 15906 bp in length
* 142209 142308: gap of 100 bp
* 142309 159777: contig of 17469 bp in length
* 159778 159877: gap of 100 bp
* 159878 178832: contig of 18855 bp in length
* 178833 178932: gap of 100 bp
* 178933 180700: contig of 1768 bp in length.

```

FEATURES

```

Source
1. 180700
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP13-506NA"
/clone_1lb="RPCT-13 Human Female BAC 481-768"
1. 951
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
1052. 2064
/note="assembly-fragment"
2165. 3416
/note="assembly-fragment"
3517. 4816
/note="assembly-fragment"
4917. 6297
/note="assembly-fragment"
6398. 9144
/note="assembly-fragment"
9245. 11578
/note="assembly-fragment"
11679. 30905
/note="assembly-fragment"
31006. 32606
/note="assembly-fragment"
32707. 34256
/note="assembly-fragment"

```

```

misc_feature 34357. 37618
/note="assembly-fragment"
misc_feature 37719. 39131
/note="assembly-fragment"
misc_feature 39232. 41291
/note="assembly-fragment"
misc_feature 41392. 43172
/note="assembly-fragment"
misc_feature 43273. 45886
/note="assembly-fragment"
misc_feature 45987. 49186
/note="assembly-fragment"
misc_feature 49287. 52781
/note="assembly-fragment"
misc_feature 52882. 57288
/note="assembly-fragment"
misc_feature 57389. 62317
/note="assembly-fragment"
misc_feature 62418. 65388
/note="assembly-fragment"
misc_feature 65489. 70061
/note="assembly-fragment"
misc_feature 70162. 75771
/note="assembly-fragment"
misc_feature 75872. 83423
/note="assembly-fragment"
misc_feature 83524. 89716
/note="assembly-fragment"
misc_feature 89817. 97026
/note="assembly-fragment"
misc_feature 97127. 104654
/note="assembly-fragment"
misc_feature 104755. 114340
/note="assembly-fragment"
misc_feature 11441. 126202
/note="assembly-fragment"
misc_feature 126303. 142208
/note="assembly-fragment"
misc_feature 142309. 159777

```

```

Query Match 64.0%; Score 16; DB 76; Length 180700;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 aggtgatgcgcagca 18
Db 25268 AGGTGATCGCCAGCA 25253

```

```

RESULT 35
AC007809
LOCUS
DEFINITION
AC007809
ACCESSION
VERSION
AC007809.8 GI:13122709
KEYWORDS
SOURCE
ORGANISM
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 183439)
Celniker, S.F., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Bantz, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierera, S., Frisoe, E., Gallego, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibeagwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,

```

Phononayong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Sytkas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, C.J.
Sequencing of Drosophila chromosome 3R, region 88c-88c
Unpublished
2 (bases 1 to 183439)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Buttenhoff, C., Chanpe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomoto, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Sytkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zierman, L.L. and Rubin, G.M.
Direct Submission
Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 24, 2001 this sequence version replaced g1:6623906.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
source
1..183439
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="88c-88c"
/clone="BACR45M03 (D718)"
/clone_11b="RPT-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"
BASE COUNT 53300 a 38849 c 38854 g 52436 t
ORIGIN

Query Match 64.0%; Score 16; DB 4; Length 183439;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcgcagcattcgat 25
|||||
Db 66707 TCGCAGCATTCGAT 66722

RESULT 36
AE003705
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AE003705 232744 bp DNA INV 05-OCT-2000
Drosophila melanogaster genomic scaffold 142000013386035 section 30 of 105, complete sequence.
AE003705 AE002708
HTG. GI:7299886
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 232744)
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brannon, R.C., Rogers, J.H., Blake, R.G., Chanpe, M., Pfeiffer, B.D., Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor

Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Baller, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bereman, B.P., Bhandari, D., Bolshakov, S., Borroa, D., Botchan, M.R., Bouck, J., Brokstein, P., Brotler, P., Butts, K.C., Busam, D.A., Butler, H., Cadieu, E., Cadenhead, C., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Meyers, A.D., Dew, I., Dietz, S.M., Dodson, K., Dunn, P., Durbin, K.J., Evans, J., Ferraz, C., Ferraz, C., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelpi, J., Gilbert, W., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ideker, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Palazzolo, M., Pittman, G.S., Pan, S., Nusskern, D.R., Pacled, J.M., Palazzolo, M., Reinert, K., Remington, K., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Siden-Kiamos, I., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Sytkas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Wortley, K.C., Wu, D., Yang, S., Yao, Q., Zhao, Q., Zheng, L., Zheng, X.H., Zaveri, J.S., Zhan, M., Zhang, G., Zhou, X., Zhu, X., Smith, H.O., Zhong, F.N., Zhong, W., Zhou, X., Zhu, X., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, C.J.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)

FEATURES
source
1..232744
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
/complement(join(29274..29439,29524..30038,30107..>30980))
/gene="CG8464"
/product="CT24781"
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0008233"
/complement(<29274..>30980)
/gene="CG8464"
/map="88c4-88c6"
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0008233"
complement(join(29274..29439,29524..30038,30107..30694))
/gene="CG8464"
/note="CG8464 gene product"
/codon_start=1
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0008233"
/protein_id="AAFP55062.1"
/db_xref="GI:7299887"
/translation="MALGSRHLEVFIERCIASPIVSHANRRSSOLATIREGPNNS
GNSGOYQNGEKEKGRRLVFEVPEELGAVNSAIIQREDLPTPTAASKMGRRD
FNEIDAVVACDSVAVYIEIKDTRHEDFSCPTIYASGSCFIIENGLITTAHYI
NKHFTWQVRLSDGRTFPATIEDVDQSDLTALRIQVNNLSVMLGSSSTIRSENV
ALGSPALAKNTYAGVISTORASOELGRNDINYLQTDALITFGNSGGLVNLDE
ATGVNSMTAGISFAIPIDYKVFLEAAEKRRKGSAYKGYVVKRMYGLTMLTLP
DILFEKSSQMPNSLTHGVYVMKVTIVVSPHSGGLPGDVIHINKKEIKNSDVI
DALADSKTLDIYILNGVKQKHVITTPBDP"
join(<31212..31346,31407..31691,31752..>31922)
/gene="CG3351"

CDs
mRNA
gene

```

/product="CT11255"
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
<31212..>31922
/gene="CG3351"
/map="88C4-88C6"
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
join(31212..31346,31407..31691,31752..31922)
/gene="CG3351"
/note="CG3351 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
/protein_id="AAFS5063.1"
/db_xref="GI:7299888"
/translation="MSKAKGKLSLKTYEVRVTHSKLTNTIPAGMAAGPLGMLG
ORAINIAPCKDINAKTAEMKEGVLPCKRISVNSDRSYDLAIHHPATFPLKQAGTIO
RGTPMPKEVAGMTLKLHLEYIAIKIDPPVNLITMOOCMLISTARTCGIKVIRE
IDPAVAGEFLERKLIIEOORRELOKREAKMLRTG"
complement(join(32136..32708,32770..>32889))
/gene="CG8461"
/product="CT24741"
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
complement(<32136..>32889)
/gene="CG8461"
/map="88C4-88C6"
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
complement(join(32136..32708,32770..32835))
/gene="CG8461"
/note="CG8461 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
/protein_id="AAFS5064.1"
/db_xref="GI:7299889"
/translation="WSSPSSSSASEHSDNFTONAIREDLKGMSFEELMKLEEL
GAKVREAVLGSRSRQPKPAKTDLKLNRNPREMTTRQVFLGAEHREKRVV
RLDRPDEKSGTVAETFPKKNQVFSOIRAKVEGOLKLDVEDEAEKHHIKDTMO
RLINKEVEDKKHKKOKOLKKEKRSATLKGOPHYLRKSTFLINSEIKF"
complement(join(<34919..35540,35598..35757,35894..36047,
36334..36679,36886..37166,37233..37324,37381..37624,
37694..37981,38058..38138,42442..42707,42765..43010,
44497..44761,48465..>49118))
/gene="CG8279"
/product="CT24224"
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
complement(<34919..>49118)
/gene="CG8279"
/map="88C5-88C6"
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
complement(join(34919..35540,35598..35757,35894..36047,
36334..36679,36886..37166,37233..37324,37381..37624,
37694..37981,38058..38138,42442..42707,42765..43010,
44497..44761,48465..49118))
/gene="CG8279"
/note="CG8279 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
/protein_id="AAFS5066.2"
/db_xref="GI:10726529"
/translation="MHGVPVRSSSSSNMNTDVSPAGGASPVEMSTSSSATNSAS
SSKRLNGANKTAISTAGGVTPGAVPGSGAIPASSSSGVNLEHHHRSNNRP
AVNRSSETKLMTGSSSPSSPOSTOISTOTISODRIAKASTTASOODVVEA
RLPEEKPEAEKMLERAPPEALSRQFTENRKHRSVTSDFQOMMASPTVVO
KSPRSLSNSASSLPECKRLHLDLDELFWELIDVANEIDLCHKRIILVNOILT
HABRGSLFLAKGTPTKNVLVAKIFDVTQATKDAVITRASAEIILIPGIGIAGVAVQ
TKOMINIKENAYKQARPNCIEDLKTGYKTNAIICMPICNVEGDIIGVAQITNKTCNMC

```

```

FEATURES
source
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AC017740
AC017740.1
GI:6554259
HTG: HTGS.PHASE2.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 237119)
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10211941 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1..237119
/organism="Drosophila melanogaster"
Query Match
Best Local Similarity 64.0%; Score 16; DB 5; Length 232744;
Matches 16; Conservativity 100.0%; Pred. No. 34;
QY 10 tcgccagcatcgcattcgat 25
Db 106438 TCGCCAGCATTCGANT 106453
FDEHDELFRRYLTPFGISGTONOLREMSVOEYRRNOILLNARSIFEBONNLECLVT
KIMEARELLKRCERGSYFLVDLDDCCASHLEKIIEKPNOPATRAIKSADSEEEKMRN
RFLYFLFELGGEYQANVSRPSVELSSFTLQI AOFVATGQTVNICVIEVRHNO
IRADELIDSTQALICMPEINMAOKRVIGVQOLINKANGVFTDSADSIIEAFYIFGGLG
IHTOMYENACKILMAKOKVALIELCSYHAGVADODPEKLIQOVYIAEAEYNLYSPFED
FEVLVDCTCRAVLRFMOCNLVSOFOIPDVLCRWLTVRKRVKRVKHHNRHNLNVA
OTFEMALCTKMERFMDLEITIGLIVAGICHLDRHGRGNNAROTYESPLAILYTTST
MEHHHDPHFLAFLADRLAARFALDELARLELDLYBELVDPITLITLITLITLITR
TILFPEPELDEPDDVDVDSVYVVLVLSVLVLSALTWATYKTIERTKT
RDNPMILALSPEDYRSVMKTVESAILISTDLAMVFKRNAFLIYENGSEFDMGEEKD
LLEGMMATACVDSIAIAKPMWVOHKAVALADEFPQGDLEKLOLNTQVAMADREKRD
ELRKMOVQIDVLCPLLYRLCDTFEPMTIPLYEGTLENRNMODLAKEVGLMIDH
DITDKVEERFAACADEIKDIEFTVTLINCNOOSOGHSESDSHTEPHOSSGRLSMKRT
GALGKAVRSKLSKTLIYNSMDGSKPKTSIKLIESHVEDMDKSPSPQPASSMKGR
MSASSTSSAGGQVMDKSKRSKRLCALL"
join(40359..40547,40526..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..>42262)
/gene="Cyp313a1"
/product="CT11271"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
<40359..>42262
/gene="Cyp313a1"
/note="CG3360"
/map="88C5-88C6"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
join(40391..40547,40526..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..42262)
/gene="Cyp313a1"
/note="Cyp313a1 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
/protein_id="AAFS5065.2"
/db_xref="GI:10726530"

```



```

/protein_id="AAF5778.1"
/db_xref="GI:7290318"
/translation="MPTWRMHRSSSSAAVEESRCRRGVGVGDANKENGVHP
MSSPFGNALIALDLINSHGKSPQRSFSGSPROATPOLAALRICTPTGCAVAL
BDSQLSSRMGDTLDRLMDLITESARKEVRCIKTLPGACATTTTLADNMASTTS
VHEMVEVPTPLKROVRRKRNPKHTTTINQOSHOKRIEPIQLVPTKRCILSFSS
SASDLDEDOOVAKRSSIASPTTPSHCTTTTSSISSSSSSNGADMAQSORSID
VSIADAKEOLANVHVRCDLORSHGSGMSINAVKVALSGAOPPGVGGHSGSGS
MSSGQRTAAVHRHSGRPFDRENFQIISGEETAGQYLQAVHRRHLNSVPTGSE
WAKYKSGTERKTARLAKLEMAENAKRESEFGSTFPLNELVHPSGVSASYKHL
AOCAPPTSRHSQPKANAGODOKETDAADODKODDSPAEMAALVTVTPKPVATG
SGSGSNSAAMALNDEVISIDSMSKEDPLPOOPOMKLSKALHODADENFLR
FLIEDPPADGANSTTQAOATGSSASANSNNHINNGISGGRSLMNSGSS
SVGAVROOQORPTMTKRLTREREGRFSTVWHPPEVEIEAGLSADRGILPG
DYVIFVKHNVMPEDVYANLIRSQSSSLTLEIFRSAGATTTITDPLGQNNVHS
TRICAAVGLSSEHTTATTAATTVMSLQRTTSTRLOPLANSMPATACGGTTSI
EAAKRLHLPOVTFKESIVPVTDNRREFFLQILISEDNFTALHFGVRFPOPIGR
KDLISPDHRTLFQNLDELRLAIEDILIELSSDODODEPOMNFASRYVLSHTTICA
AYKKGNGIKRACDVLNKSROGSEFIATFEPVPRKRPDLMTHTPHLOHFEIL
KIMOLAGNCHVDIEEHKNSSTVINELQAAVREITYSSGLMEPLGGRPLTLQDLES
RMVPTKCKPPTLVAGROMTFGDLSRVGRSVKPYWLLFSDIIVFAVSRDRLFI
TEEPPIANVVDSCFPHRKKTTFEFLTVDPNGKLASPTGYCAPDLTRTPKGAARKS
LILRASPILKAWNLDLQKQIFLVNAAAGSTPLSSPLDPLVTLVPLSDIGLTTA
SGMSKRLPSLDSIHLKQOQOQVRLFRSRKLDGSDYENLQKLAASAVRHHCATTS
ATQTSHPHPPLOCTTSLPSRTNSPARNOHPOGCVGGVSGVGGVSGFGCANNSS
CLSSASITQTOVHTQILNHSOTHSOTQTSVILSSCKCLTVIPVTSPPVQHA
SOKLEFSLSSVGRSFLIDSGCVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGV
ATSTGKGNPIPLAERGSMDLELDLQHEVNDLPSTDTDVEECIFLGEDRGTGE
GHGHDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD
complement(<53568..>54629)
/gene="EG:BACH7M4.4"
/product="CT33234"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/complement(<53568..>54629)
/gene="EG:BACH7M4.4"
/note="CG12496"
/map="286-3A1"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/complement(53568..54629)
/gene="EG:BACH7M4.4"
/note="EG:BACH7M4.4 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/protein_id="AAF5779.1"
/db_xref="GI:7290319"
/translation="MLKTTASQRLNLEIPIVHDODANOPORRRLORRORRATANK
ENVPVTYTTSPVITRLNRSPLVLAFLSDIRNVTPEAVVRSQAPORVOSVYATL
PFAEDYSPNGLPSGOHLALIGLAFLDPELQPRYIVGSGAGVAKLKORLNDPFAAT
LEPEDETAVEAPEPTPKVEVRSTPOPASTOMQPDRLIDALILDSACANMSKK
PRSTFNLRRRTLVKQOMESNCPQEVESLSPADPDASPLSGCLVPLPHNIMPTPE
PASPLAIPHLVLVOLAIPIVAPKAPCDNFTFCLTPKALKRTRKEIVAKESPTKRYK
VDERNFEVGLALPSSITYV"
complement(<70555..71849,71925..>71976))
/gene="gt"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT23972"
/db_xref="FLYBASE:FBan0007952"
/db_xref="FLYBASE:FBgn0001150"
/complement(<70555..>71976)
/gene="gt"
/note="CG7952"
/map="3A1-3A1"
/db_xref="FLYBASE:FBan0007952"

```

Query Match 64.0%; Score 16; DB 4; Length 289090;
Best Local Similarity 100.0%; Pred. No. 34;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 gtfcatcgccagcatt 20
Db 13743 gtfcatcgccagcatt 13758
RESULT 39
NPTPLG/c
LOCUS NPTPLG 1389 bp DNA PLN 29-JUL-1996
DEFINITION N. tabacum gene for pectate lyase.
ACCESSION X61102
VERSION X61102.1 GI:19981
KEYWORDS pectate lyase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
1; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 1389)
Lonsdale, D.M.
REFERENCE 2
Submitted (25-JUL-1991) D.M. Lonsdale, Cambridge Laboratory, Centre
for Plant Science Research, Colney Lane, Norwich, Norfolk NR4 7UJ,
UK
2 (bases 1 to 1389)
Rogers, H.J., Harvey, A. and Lonsdale, D.M.
Isolation and characterization of a tobacco gene with homology to
pectate lyase which is specifically expressed during
microsporogenesis
Plant Mol. Biol. 20 (3), 493-502 (1992)
See X61101 for N. tabacum pectate lyase mRNA (partial).
LOCATION/Qualifiers
1. 1389
/organism="Nicotiana tabacum"
/strain="Samsun"
/db_xref="taxon:4097"
/dev_stage="mature pollen"
/clone_1b="lambda-EMBL3"
/clone="G10"
<1..577
/number=1
join(1..577,676..976,1074..1389)
CDS
prim_transcript <1..>1389
exon
join(1..577,676..976,1074..1389)
/codon_start=1
/product="pectate lyase"
/protein_id="CAA3414.1"
/db_xref="GI:19982"
/db_xref="SWISS-PROT:PA0972"
/translation="MDVYRIRISVFLLVLTTPAALTATNIPRQLSNKKYKGPCRA
ENALDKRCRCPMAENRQKMDALDFGSAIGKLGIRYYVTNDSDDDVDPVPGT
LRYGVLOKEPLWILTEGKNMFKLSRELIVTSNKTIDGRFNHIIQMGAGIKIOSASNI
LISNLRHNIVPPGGILRSEDEHVIGSDESDGISFSHDIMWDHISMRATDGL
IDAAVASTNITISNCHTDEKWLFCANDHYVLDDKMTLLAYNHEGKRLLDORPRC
RFGFHHVNNIDYTHMERYAIGGSGAFTIISQGRFPADELLIKETVYREKRTASVAV
MKMTWTISDDGDMMENATFTPPSDQNLIDKIDHLNLIKPEPSKVGIIITKFGALSTCY
KGRPC"
join(<1..577,676..976,1074..>1389)
578..675
/number=1
676..976
/number=2
977..1073
/number=2
1074..>1389
/number=3
BASE COUNT 427 a 305 g 387 t
ORIGIN

```

Query Match 60.0%; Score 15; DB 14; Length 1389;
Best Local Similarity 100.0%; Pred. No. 2;le102;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatgccagcattt 21
 |||||
 Db 162 GCATGCCAGCATTT 148

RESULT 40

LOCUS NTPECL1 1713 bp DNA PLN 29-JUL-1993
 DEFINITION N. tabacum gene for pectate lyase.
 X67158.547561
 VERSION X67158.1 GI:19907
 KEYWORDS pectate lyase.
 SOURCE common tobacco.
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridicophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
 1; Solanales; Solanaceae; Nicotiana.

REFERENCE

1 (bases 1 to 1713)
 Lonsdale, D.M.
 AUTHORS Lonsdale, D.M.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUL-1992) D.M. Lonsdale, Cambridge Laboratory, Centre
 for Plant Science Research, Colney Lane, Norwich Norfolk NR4 7UJ, UK

REFERENCE

2 (bases 1 to 1713)
 Rogers, H.J., Harvey, A. and Lonsdale, D.M.
 AUTHORS Rogers, H.J., Harvey, A. and Lonsdale, D.M.
 TITLE Isolation and characterization of a tobacco gene with homology to
 pectate lyase which is specifically expressed during
 microsporogenesis

JOURNAL

Plant Mol. Biol. 20 (3), 493-502 (1992)

MEDLINE

93043039

FEATURES

source Location/Qualifiers
 1..1713
 /organism="Nicotiana tabacum"
 /strain="Samsun"
 /db_xref="taxon:4097"
 /gene="NtPECL1"
 /dev_stage="mature pollen"
 /clone_lib="lambda EMBL3"
 /clone="G10"
 108..1713
 prim_transcript
 108..901
 exon
 join(108..901,1000..1300,1398..1713)
 mRNA
 325..399
 translt_peptide
 join(325..901,1000..1300,1398..1713)
 CDS
 /product="pectate lyase"
 /codon_start=1
 /protein_id="CAA47630.1"
 /db_xref="GI:19908"
 /db_xref="SWISS-PROT:P40972"
 /translation="MDVYRIISVFLVLTLPFAALTTPATNIPRQLSNKKYKGCRA
 ENALDKCMRCDDPMWENRQKMDACAFSGSNAIGKLRIVYVTDNSDDVDVDPGPT
 LRYGVIOKEPILWITFGKNNKIKSLRELYTSNKTIDRGFNHIONGAGIKIOSASNI
 ITSNLRHNIVTPPGCLRESDEHVGILSGSDSGISSTSHDIWIDHISMRAIDGL
 IDVAASNTNITISNCHETDHEKYMALGANDHYLQDKMTITLAINHEGKRIDQRPKC
 RCFEFLVNDYTHWERYALGSSGATITISQNRFAEDLIVKEVYREKLIVASVAE
 WMKWTISDGDMDENGATFTPSGDQNLIDKLIDHLNLIKPESSKVGILTRFSGALSCV
 KGRPC"

mat_peptide
 join(400..901,1000..1300,1398..1710)
 /product="pectate lyase"
 902..999

intron

exon

intron

exon

BASE COUNT

546 a 353 g 476 t

ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 1713;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatgccagcattt 21
 |||||
 Db 486 GCATGCCAGCATTT 472

RESULT 41

LOCUS HSSYNTRO5 1907 bp mRNA PRI 27-MAY-2000
 DEFINITION Homo sapiens mRNA for syntrophin 5.
 X67158.547561
 VERSION AJ003029.1 GI:8247276
 KEYWORDS syntrophin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 1907)
 Piluso, G., Belisio, A., Puca, G.A. and Nigro, V.
 AUTHORS Piluso, G., Belisio, A., Puca, G.A. and Nigro, V.
 TITLE Identification of a novel syntrophin-like cDNA
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1907)
 Nigro, V.

AUTHORS

Direct Submission
 Submitted (11-NOV-1997) Nigro V., Istituto di Patologia Generale ed
 Oncologia, Seconda Università degli Studi di Napoli, Larghetto S.
 Aniello a Caponapoli 2, 80138, ITALY

JOURNAL

Unpublished

FEATURES

source Location/Qualifiers
 1..1907
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p25"
 132..1751
 /codon_start=1
 /product="syntrophin 5"
 /protein_id="CAB62969.1"
 /db_xref="GI:8247277"
 /db_xref="GI:8247277"
 /translation="MGTGGPPPPAASRGQCLVLPARTKTTTALLIPESENNADIR
 LKLTKEVLTIOKODVVCVGSQGRNRRTVLRQPVGGLSTIGSGSEHNPVVISK
 IFEDQADQGMLEFVDALVQNGIHVENATHEEVHLIRNAGDEVITTVYLRAPL
 FLKPLPGSPGSDHSSGASPLFDGILNNGSTTAPSSSPFIADPRYERWLD
 TLSVPLSMARISRYKAEKIRRNNAFEVILGVSGLIRFTADGDMLRAVSANI
 RELTLDNMKANKCGSPSDVYVMGVNKKLOGAASSOTFRPKPLALGSPSEYSP
 PVSTFDVMRAERTYHLCEVLFVHKFWLTEDCWLADLIGLQDFDFDQRPYCSIV
 AGHGSKSHVEVELGSELAMERKSFQATPEVQRTGSRVYKSWGEMLCFTVDALG
 FTCESEKTKVNLIRFKFESOLKSSDGGKTRVLLRFQNDLTKOIEKKLEFDLRAVLH
 CHISFTAAKVASVDPGPMDSQSLARKYVSS"

BASE COUNT

474 a 476 c 537 g 420 t

ORIGIN

Query Match 60.0%; Score 15; DB 93; Length 1907;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcattt 22
 |||||

Db 570 CATGCCAGCATTTT 556

RESULT 42

LOCUS AB048915 2616 bp mRNA PRI 20-SEP-2000
 DEFINITION Macaca fascicularis brain cDNA, clone:QnpA-17839.
 X67158.547561
 VERSION AB048915.1 GI:10241989
 KEYWORDS f1s (full insert sequence).
 SOURCE Macaca fascicularis adult male brain parietal lobe cDNA to mRNA,

ORGANISM
clone_lib:macaque brain cDNA library Qnpa clone:Qnpa-17839.
Macaque fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.

REFERENCE
AUTHORS
1 (sites)
Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished (2000)

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 2616)
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
Direct Submission
Submitted (14-SEP-2000) to the DDBJ/EMBL/Genbank databases.
Katsuyuki Hashimoto, National Institute of Infectious Diseases,
Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku,
Tokyo 162-8640, Japan (E-mail:hashi@nih.go.jp,
URL:http://www.nih.go.jp/yoken/genbank//,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab Host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
were constructed by Sugano et al.(University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
{ 5' end primer [CGACCTGCACCTCGACGACA] };
{ 3' end primer [CTTTCGCTCTTAAAGCTGCG] }.

FEATURES
source
Location/Qualifiers
1..2616
/organism="Macaque fascicularis"
/db_xref="taxon:9541"
/clone_lib="Qnpa-17839"
/clone_lib="macaque brain cDNA library Qnpa"
/dev_stage="adult"
/sex="male"
/tissue_type="brain parietal lobe"

BASE COUNT
ORIGIN
739 a 527 c 572 g 778 t

Query Match
Best Local Similarity 60.0%; Score 15; DB 85; Length 2616;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccacgacattc 22
|||||
Db 1743 CATGCCACGACATTTC 1729

RESULT 43
AF274670 3369 bp DNA PLN 13-AUG-2000
LOCUS Tetracentron sinense 26S ribosomal RNA gene, complete sequence.
DEFINITION AF274670
ACCESSION AF274670
VERSION AF274670.1 GI:9799468
KEYWORDS
SOURCE Tetracentron sinense.
ORGANISM Tetracentron sinense
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Trochodendraceae; Tetracentroideae;
Tetracentron
1 (bases 1 to 3369)
Fishbein,M., Hilsch-Jetter,C., Solits,D.E. and Huford,L.
Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
Rapid, Ancient Radiation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 3369)
Fishbein,M., Hilsch-Jetter,C., Solits,D.E. and Huford,L.
Direct Submission
Submitted (30-MAY-2000) School of Biological Sciences, Washington
State University, Pullman, WA 99164-4236, USA

FEATURES
source
Location/Qualifiers
1..3369
/organism="Tetracentron sinense"
/specimen_voucher="Qiu 94166 (IND)"
/db_xref="taxon:13715"
1..3369
/product="26S ribosomal RNA"

BASE COUNT
ORIGIN
806 a 824 c 1049 g 689 t 1 others

Query Match
Best Local Similarity 60.0%; Score 15; DB 13; Length 3369;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcattgcacgc 17
|||||
Db 441 AGGTGCATGCCACGC 427

RESULT 44
AF274671 3369 bp DNA PLN 13-AUG-2000
LOCUS Trochodendron aralioides 26S ribosomal RNA gene, complete sequence.
DEFINITION AF274671
ACCESSION AF274671
VERSION AF274671.1 GI:9799469
KEYWORDS
SOURCE Trochodendron aralioides.
ORGANISM Trochodendron aralioides
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Trochodendraceae; Trochodendroideae;
Trochodendron
1 (bases 1 to 3369)
Fishbein,M., Hilsch-Jetter,C., Solits,D.E. and Huford,L.
Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
Rapid, Ancient Radiation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 3369)
Fishbein,M., Hilsch-Jetter,C., Solits,D.E. and Huford,L.
Direct Submission
Submitted (30-MAY-2000) School of Biological Sciences, Washington
State University, Pullman, WA 99164-4236, USA

FEATURES
source
Location/Qualifiers
1..3369
/organism="Trochodendron aralioides"
/specimen_voucher="Qiu 94157 (IND)"
/db_xref="taxon:4407"
1..3369
/product="26S ribosomal RNA"

BASE COUNT
ORIGIN
801 a 816 c 1060 g 690 t 2 others

Query Match
Best Local Similarity 60.0%; Score 15; DB 13; Length 3369;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcattgcacgc 17
|||||
Db 441 AGGTGCATGCCACGC 427

RESULT 45
PSE1AAMH 3677 bp DNA BCT 26-APR-1993
LOCUS PSE1AAMH/c
DEFINITION P. syringae tryptophan 2-monooxygenase (1aam) and indoleacetamide
hydrolase (1aah) genes, complete cds.

ACCESSION M1035 M35690
 VERSION M1035.1 GI:151289
 KEYWORDS indoleacetamide hydrolase; indoleacetic acid; tryptophan
 2-monooxygenase; tryptophan-2-oxyreductase.
 SOURCE *P. syringae* savastanoi (Strain EM2009) DNA, clones pLOC2, pCP3, and pCP12.
 ORGANISM *Pseudomonas syringae*
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 202 to 3677)
 AUTHORS Yamada, T., Palm, C.J., Brooks, B. and Kosuge, T.
 TITLE Nucleotide sequences of the *Pseudomonas savastanoi* indoleacetic acid genes show homology with *Agrobacterium tumefaciens* T-DNA
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82, 6522-6526 (1985)
 REFERENCE 2 (bases 1 to 312)
 AUTHORS Gaffney, T.D., da Costa e Silva, O., Yamada, T. and Kosuge, T.
 TITLE Indoleacetic acid operon of *Pseudomonas syringae* subsp. *savastanoi*: Transcription analysis and promoter identification
 JOURNAL J. Bacteriol. 172, 5593-5601 (1990)
 COMMENT The *iaaH* and *iaaM* genes are part of an operon that is borne on a plasmid, pI_{IAA}, in *Agrobacterium* strains of the pathogen. There appears to be no genetic transformation of host tissue by *P. savastanoi*. No *Pribnow* boxes were found, though potential ribosome binding sites are located at positions 435-439 and 2164-2168. Expression of *iaaH* depends upon the promoter for *iaaM*.
 Draft entry and computer-readable sequence for [1] kindly submitted by T. Yamada, 14-JAN-1986.
 Draft entry and computer-readable sequence for [J. Bacteriol. (1990) in press] kindly submitted by T.D. Gaffney, 28-JUN-1990.
 FEATURES
 source
 1..3677
 /organism="Pseudomonas syringae"
 /sub_species="savastanoi"
 /db_xref="taxon:317"
 19..47
 /note="indoleacetic acid operon promoter"
 54..>3537
 /note="indoleacetic acid mRNA"
 82..312
 /note="promoter proximal ORF"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25851.1"
 /db_xref="GI:151290"
 /translation="MPGRTPANWLDKAMVAVARIRQRKPOAAVPAALLQAPHAMFEVW
 STTKPTTECHCPATPYTQASAMASAVVAY"
 454..2127
 /note="tryptophan 2-monooxygenase (EC 1.13.12.3)"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25852.1"
 /db_xref="GI:151291"
 /translation="MYDHNPSIDILYGPFLKCKEMTGIGSYASAGTPPRAVAIV
 GAGISGLVATTELIRAGCVKDVVYIESRDRIGRWYQVDPQRRPIAEMGAMREPPS
 AGLPHYIAKFGISTSTTPDPGVVDTELYRGRYHMPAGKPELFRVYEGWQSL
 LSEGYLLEGGSLVAPLDITAMLSKGRLEAAIAWQGLNVFDCSYNAIVC1FTGRH
 PGQDWARPEDELEGLSGISGFLPVOAGFTEILRMVINGVQSDRLIPDGISS
 LAARLADQSEFDGALDRVCFSVGRISREAEKIIITQAGQREVDRIYVTSNNRM
 OMHCLTDESEFLSRDVARAVRETLTGSSKLPILTRKFWIKNKLPTIIOSDGLVYG
 VYLDYQDPEPEGHGVLLSTYMEDDAOKMLPMPDKKTCQVLYVDLIAIHPTFASYL
 LPYDGYERYVLIHDMULTDPHSGAFKLNYPGEDYVSORLFPQPTANSFNKDTGLYL
 AGSCSFAGGWMIEGAVQTALNSACAVLRSTGSLSKGNPLDCINASYR"
 2170..3537
 /note="indoleacetamide hydrolase"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25853.1"
 /db_xref="GI:151292"
 /translation="MHEITLESLLQCALADGEIAAELKERALDTEARLARLNFIRE
 GNAVSGEGADHAMKCTPLMGKPVFSKDNICVAGLPITAGTGMGSEFVSDDDAIVSQ
 LRLAGVAVAGKNNMHELSPGVTSINPHMCTVGNPVAFGYAGGSSGSGSAAVAAGIYP

BASE COUNT 813 a 917 c 1046 g 901 t
 ORIGIN 1 bp upstream of EcoRI site.

Query Match 60.0%; Score 15; DB 3; Length 3677;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatgcacgcatc 21
 |||
 DB 1729 GCATGCCGAGCATTT 1715

Search completed: October 9, 2001, 15:46:36
 Job time: 14782 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:25 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25

Sequence: 1 gccaggtgcatcgccagcatlccgat 25

Scoring table: OLIGO.MUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 9

Total number of hits satisfying chosen parameters: 28031

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*

1: /SIDSL/gcgdata/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E. coli bio B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAO62386	Biotin-biosynthesi
5	16	64.0	1498	21 AAC51224	Arabidopsis thalia
6	16	64.0	1500	21 AAC36258	Arabidopsis thalia
7	15	60.0	2597	21 AAC93313	Human secreted pro
8	15	60.0	5773	21 AAC76014	Human ORFX ORF1569
9	14	56.0	549	21 AAC93854	Cat flea hindgut a
10	14	56.0	719	19 AAV20504	Conus geographus c
11	14	56.0	719	19 AAV17130	Conus geographus c

12	14	56.0	1190	21 AAC98831	Human pancreatic c
13	14	56.0	1344	21 AA256381	Escherichia coli f
14	14	56.0	2017	21 AA659762	Human ovarian carc
15	14	56.0	2193	17 AAT30684	Kaposi's sarcoma a
16	14	56.0	2193	17 AAT16814	Kaposi's sarcoma a
17	14	56.0	2193	20 AA220914	Nucleotide sequenc
18	14	56.0	3593	21 AA239011	Mouse Ees2 coding r
19	14	56.0	4189	21 AA249334	Mouse Ees2 coding r
20	14	56.0	4313	14 AA038950	Mouse Ees2 full le
21	14	56.0	4625	21 AA339010	Mouse Ees2L coding
22	14	56.0	4975	21 AA239027	R. prowazekii S-la
23	14	56.0	5319	19 AA355235	Amycolatopsis medi
24	14	56.0	5676	19 AAV21186	Mouse Ees2L CDNA s
25	14	56.0	6014	21 AA339026	Kaposi's sarcoma a
26	14	56.0	20710	17 AAT16806	KSHV LTR DNA (nucl
27	14	56.0	20710	17 AAT16806	Amycolatopsis medi
28	14	56.0	35100	20 AAV73803	KSHV long unique c
29	14	56.0	53789	19 AAV21187	Oligonucleotide #3
30	14	56.0	137507	19 AAV19941	Human clone c92838
31	13	52.0	22	19 AAV11150	Human secreted exp
32	13	52.0	51	21 AA76686	Human gene express
33	13	52.0	51	21 AA76687	Human gene express
34	13	52.0	286	21 AA45502	Human gene express
35	13	52.0	300	20 AA214744	Plant microsatelli
36	13	52.0	300	20 AA214726	Hepatitis C virus
37	13	52.0	322	21 AAA31174	Secreted protein E
38	13	52.0	447	17 AAT7962	Fusarium venenatum
39	13	52.0	538	20 AA240788	Fusarium venenatum
40	13	52.0	629	21 AAF09259	Fusarium venenatum
41	13	52.0	632	21 AAF09268	Eucalyptus grandis
42	13	52.0	646	21 AAC55942	Human gene express
43	13	52.0	714	20 AA217005	Human breast cancer
44	13	52.0	725	22 AAF22436	Human gene express
45	13	52.0	749	20 AA216470	Human gene express

ALIGNMENTS

RESULT 1	
AA01303	standard: DNA; 1041 BP.
AA01303	
AA01303	
12-APR-1999	(first entry)
E. coli biotin synthetase (Biot) coding sequence.	
DAP aminotransferase; diaminopelargonic acid; transgenic plant;	
biotin synthase; biotin production; vitamin H; Biot; ss.	
Escherichia coli.	
US5869719-A.	
09-FEB-1999.	
30-APR-1997.	97US-0846338.
30-APR-1997.	97US-0846338.
08-MAR-1995.	95US-0401068.
(NOVS) NOVARTIS FINANCE CORP.	
Patton DA.	
PI	
WPI, 1999-152902/13.	
P-PSDB; AAV73906.	
Transgenic plants with high biotin levels - transformed with DNA	
encoding di- amino- pelargonic acid amino- transferase or biotin	
synthase	

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (Biot). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 Db 84 gcagggtgcatgcgcagcattcgat 108

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

FH Key Location/Qualifiers

FT CDS 24..1064

PN GB2216530-A.

PD 11-OCT-1989.

PE 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pre- control sequences for expression in S.cerevisiae
 CC are plasmids pMA51, pMA56c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 Db 107 gcagggtgcatgcgcagcattcgat 131

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desbiobiotin; ds.

FH Key Location/Qualifiers

FT CDS 42..1082

PN JP61149091-A.

PD 07-JUL-1986.

PE 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAN60536.

PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 Db 125 gcagggtgcatgcgcagcattcgat 149

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

AC AAO62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; biob; biof; bioc; biod; bioA;
 KW promoter; plac; biotin synthase; KAPA synthase;
 KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;


```
XX      EPI033405-AZ.
PN
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123380.
PR      09-MAR-1999; 99US-0123348.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131445.
PR      04-MAY-1999; 99US-0132048.
PR      05-MAY-1999; 99US-0132407.
PR      06-MAY-1999; 99US-0132484.
PR      07-MAY-1999; 99US-0132486.
PR      11-MAY-1999; 99US-0132863.
PR      14-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135124.
PR      21-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      17-JUN-1999; 99US-0139453.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139877.
PR      22-JUN-1999; 99US-0139889.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.

PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144321.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      25-JUL-1999; 99US-0145224.
PR      27-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      06-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147302.
PR      06-AUG-1999; 99US-0147416.
PR      09-AUG-1999; 99US-0147493.
PR      09-AUG-1999; 99US-0147935.
PR      10-AUG-1999; 99US-0148171.
PR      11-AUG-1999; 99US-0148319.
PR      12-AUG-1999; 99US-0148341.
PR      13-AUG-1999; 99US-0148565.
PR      13-AUG-1999; 99US-0148684.
PR      16-AUG-1999; 99US-0149368.
PR      17-AUG-1999; 99US-0149175.
PR      18-AUG-1999; 99US-0149426.
PR      20-AUG-1999; 99US-0149722.
PR      20-AUG-1999; 99US-0149723.
PR      20-AUG-1999; 99US-0149929.
PR      23-AUG-1999; 99US-0149902.
PR      23-AUG-1999; 99US-0149910.
PR      23-AUG-1999; 99US-0150566.
PR      26-AUG-1999; 99US-0150884.
PR      27-AUG-1999; 99US-0151065.
PR      27-AUG-1999; 99US-0151066.
PR      27-AUG-1999; 99US-0151080.
PR      30-AUG-1999; 99US-0151303.
PR      31-AUG-1999; 99US-0151438.
PR      01-SEP-1999; 99US-0151930.
PR      07-SEP-1999; 99US-0152363.
PR      10-SEP-1999; 99US-0153070.
PR      13-SEP-1999; 99US-0153758.
PR      15-SEP-1999; 99US-0154018.
```

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155586.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 2; 6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agtgcacgcacgca 18
|||||
DB 84 agtgcacgcacgca 99

RESULT 6

AAC36258
ID AAC36258 standard; DNA; 1500 BP.

XX AAC36258;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126764.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142205.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

```

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match          64.0%; Score 16; DB 21; Length 1500;
Best Local Similarity 100.0%; Pred. No. 2.0;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 aggtgcatgcgcagca 18
Db 84 aggtgcatgcgcagca 99

```

```

RESULT 7
AAC93313
ID AAC93313 standard; cDNA; 2597 BP.
XX
AC AAC93313;
XX
DT 16-FEB-2001 (first entry)
XX

```

```

DE Human secreted protein cDNA sequence #4.
XX

```

```

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuc;
KW valiant; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW caridiat; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX

```

```

OS Homo sapiens.
XX
PN W0200058495-A1.
XX

```

```

PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000MO-US07661.
XX

```

```

PR 26-MAR-1999; 99US-0126504.
PR 07-JAN-2000; 2000US-0174847.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

```


PI Rosen CA, Ruben SM, Komatsoulis G;
 XX WPI: 2000-611720/58.
 DR P-PSDB: AAB51383.
 XX
 PT New nucleic acid molecules encoding 45 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; Pages 327-328; 410pp: English.
 XX
 CC The invention relate to the isolation of genes AAC93310-C93354 encoding
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 SQ Sequence 2597 BP; 721 A; 604 C; 534 G; 737 T; 1 other;
 OY
 Query Match 60.0%; Score 15; DB 21; Length 2597;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 gcaagtgatcgcca 15
 |||||||||
 Db 2256 gcaagtgatcgcca 2270
 RESULT 8
 AAC76014/c
 ID AAC76014 standard; cDNA: 5773 BP.
 AC AAC76014;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
 XX
 KW Human: open reading frame; ORFX: detection: cytostatic; hepatotropic;
 KW vulnerability: antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antihemetic; antithyroid;
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.

XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX WPI: 2000-602362/57.
 DR P-PSDB: AAB41805.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2354-2357; 5507pp: English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antiparkinsonian; nootropic; neuroprotective;
 CC osteoporotic; antiparkinsonian; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antihemetic;
 CC antithyroid; and antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
 OY
 Query Match 60.0%; Score 15; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 4886 CATCGCACGACATTTC 4872
 |||||||||
 OY 8 catcgacagatttc 22
 |||||||||
 Db 4886 CATCGCACGACATTTC 4872
 RESULT 9
 AAC93854/c
 ID AAC93854 standard; cDNA: 549 BP.
 AC AAC93854;
 XX
 DT 19-FEB-2001 (first entry)
 DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:349.
 XX
 KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
 KW flea infestation; vaccine; antiparasitic; therapeutic target;
 KW diagnosis; detection; ss.
 XX
 OS Ctenocephalides felis.
 XX
 PN WO200061621-A2.
 XX
 PD

PD 19-OCT-2000.
 XX
 XX 07-APR-2000; 2000WO-US09437.
 XX
 XX 09-APR-1999; 99US-0128704.
 XX
 XX (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 XX
 XX WPI: 2000-656323/63.
 DR
 XX
 PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 PT acids useful for the prevention, diagnosis and treatment of flea
 PT infestations -
 XX
 XX
 PS Claim 26; Page 361; 964pp; English.

XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
 CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 CC or head and nerve cord (HNC) tissue. The invention also relates to the
 CC encoded proteins. The invention additionally encompasses expression
 CC constructs, recombinant viruses and recombinant cells comprising the
 CC nucleic acids of the invention, recombinant production of the proteins,
 CC antibodies against the proteins, a method of identifying inhibitors of
 CC the proteins, and compositions comprising the inhibitors for
 CC administration to an animal. The nucleic acids, and the proteins they
 CC encode may be used in the prevention, treatment and diagnosis of diseases
 CC associated with flea infestations. For example, the nucleic acids may be
 CC used to produce an HMT or HNC protein according to standard recombinant
 CC DNA methodology by inserting the nucleic acids into a host cell and
 CC culturing the cell to express the protein. The HMT and HNC nucleic acids
 CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
 CC and quantitate the presence of cat flea or other homologous nucleic acid
 CC sequences in samples. They may also be used to study the expression and
 CC function of the proteins and their role in metabolism. The HMT and HNC
 CC proteins may be used as antigens in the production of specific
 CC antibodies, and in assays to identify modulators (agonists and
 CC antagonists) of HMT and/or HNC protein expression and activity. The
 CC anti-HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HMT cDNA of the invention.
 XX
 XX Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 549;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catcgccagcatll 21
 Db 421 CATGCCAGCATTT 408
 |||||

RESULT 10
 AAV20504/c
 ID AAV20504 standard; DNA; 719 BP.
 XX
 XX AAV20504;
 AC
 XX
 DT 30-JUN-1998 (first entry)
 XX
 DE Conus geographus conantokin DNA.
 XX
 KM Conantokin; predatory cone snail; treatment; neurologic disorder;
 KM psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
 KM HIV infection; ophthalmic indication; memory; learning defect;
 KM cognitive defect; ss.
 XX
 XX Conus geographus ..
 OS

XX Key Location/Qualifiers
 XX CDS 110..412
 XX /tag= a
 XX /product= conantokin
 XX
 XX MO9803541-AI.
 XX
 XX 29-JAN-1998.
 PD
 XX
 XX 21-JUL-1997; 97WO-US12618.
 PF
 XX
 XX 22-JUL-1996; 96US-0684742.
 PR
 XX
 XX (COGN-) COGNETIX INC.
 PA (UTAH) UNIV UTAH RES FOUND.
 XX
 XX Abogadie EC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;
 PI Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;
 PI Walker C, Zhou L;
 XX
 XX WPI: 1998-120694/11.
 DR P-PSDB; AAM48210.
 XX
 XX
 PT New conantokin peptide(s) - useful for e.g. treating neurologic or
 PT psychiatric disorders, or the management of pain
 XX
 XX Claim 19; Pages 79-80; 122pp; English.

XX The present sequence encodes Conus geographus conantokin, peptide
 CC derivatives of which can be used to treat neurologic and
 CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
 CC or analgesic agent. Neurologic and psychiatric disorders include
 CC epilepsy, convulsions, neurotoxic injury (associated with
 CC conditions of hypoxia, anoxia or ischemia, which typically follow
 CC stroke, cerebrovascular accident, brain or spinal cord trauma,
 CC myocardial infarct, physical trauma, drowning, suffocation,
 CC perinatal asphyxia or hypoglycaemic events), neurodegeneration
 CC associated with Alzheimer's disease, senile dementia, Amyotrophic
 CC Lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
 CC Huntington's disease, Down's Syndrome, Korsakoff's disease,
 CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
 CC dementia and neuronal damage associated with uncontrolled
 CC seizures), chemical toxicity (such as addiction, and morphine,
 CC opiate, oploid and barbiturate tolerance), pain (acute, chronic,
 CC migraine), anxiety, major depression, manic-depressive illness,
 CC obsessive-compulsive disorder, schizophrenia and mood disorders
 CC (such as bipolar disorder, unipolar depression, dysthymia and
 CC seasonal affective disorder) and dystonia (movement disorder),
 CC sleep disorder, muscle relaxation and urinary incontinence. The
 CC peptide can also be used to treat HIV infection, ophthalmic
 CC indication and memory, learning or cognitive defects.
 XX

Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 caggtgcatcgcca 15
 Db 118 CAGGTGCAATCGCCA 105
 |||||

RESULT 11
 AAV17130/c
 ID AAV17130 standard; DNA; 719 BP.
 XX
 XX AAV17130;
 AC
 XX
 DT 30-JUN-1998 (first entry)
 XX

DE Conus geographus conantokin DNA.

KW Conantokin; predatory cone snail; treatment: neurologic disorder;
 KW psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
 KW HIV infection; ophthalmic indication; memory; learning defect;
 KW cognitive defect; ss.

XX Conus geographus.

OS

XX

XX Key Location/Qualifiers

FT CDS 110..412

FT /*tag= a

FT /product= conantokin

XX

XX W09803189-A1.

XX

XX

XX 29-JAN-1998.

XX

XX 21-JUL-1997; 97WO-US12652.

XX

XX 06-DEC-1996; 96US-0762377.

XX 22-JUL-1996; 96US-0684750.

XX

XX (COGN-) COGNETIX INC.

XX

XX Layer RT, McCabe RT, Zhou L;

XX

XX WPI: 1998-120469/11.

XX P-PSDB: AAW49989.

XX

XX Use of conantokin peptide(s) - for treating disorders involving
 PT excessive excitation of nerve cells by excitatory amino acids or
 PT agonists of the N-methyl-D-aspartate receptor

XX

XX Example 4; Pages 79-80; 122pp; English.

XX

XX The present sequence encodes Conus geographus conantokin, peptide
 CC derivatives of which can be used to treat neurologic and
 CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
 CC or analgesic agent. Neurologic and psychiatric disorders include
 CC epilepsy, convulsions, neurotoxic injury (associated with
 CC conditions of hypoxia, anoxia or ischaemia, which typically follow
 CC stroke, cerebrovascular accident, brain or spinal cord trauma,
 CC myocardial infarct, physical trauma, drowning, suffocation,
 CC perinatal asphyxia or hypoglycaemic events), neurodegeneration
 CC associated with Alzheimer's disease, senile dementia, Amyotrophic
 CC lateral sclerosis, Multiple Sclerosis, Parkinson's disease,
 CC Huntington's disease, Down's Syndrome, Korsakoff's disease,
 CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
 CC dementia and neuronal damage associated with uncontrolled
 CC seizures), chemical toxicity (such as addiction, and morphine,
 CC opiates, opioid and barbiturate tolerance), pain (acute, chronic,
 CC migraine), anxiety, major depression, manic-depressive illness,
 CC obsessive-compulsive disorder, schizophrenia and mood disorders
 CC (such as bipolar disorder, unipolar depression, dysthymia and
 CC seasonal affective disorder) and dystonia (movement disorder),
 CC sleep disorder, muscle relaxation and urinary incontinence. The
 CC peptide can also be used to treat HIV infection, ophthalmic
 CC indication and memory, learning or cognitive defects.

XX

XX Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cagtgatcgccca 15
 |||||||||||||
 DB 118 CAGGTGATCGCCA 105

RESULT 12

AAC98831
 ID AAC98831 standard; CDNA; 1190 BP.

XX

XX AAC98831;

XX

XX 09-MAR-2001 (first entry)

XX

XX Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.

XX

XX

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative; ss.

XX

XX Homo sapiens.

XX

XX W0200055320-A1.

XX

XX 21-SEP-2000.

XX

XX 08-MAR-2000; 2000WO-US05989.

XX

XX 12-MAR-1999; 99US-0124270.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM;

XX

XX WPI: 2000-579444/54.

XX P-PSDB: AAB54066.

XX

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -

XX

XX Claim 1; Page 537; 1379pp; English.

XX

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX

XX Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1190;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgcatgcgcacg 17
 |||||||||||||
 DB 1141 ggtgcatgcgcacg 1154

RESULT 13

AAZ56381/c
ID AAZ56381 standard; DNA: 1344 BP.

XX AAZ56381;

XX 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX flagellin; fljC; antigen; detection; ds.

XX Escherichia coli.

XX WO9961458-A1.

XX 02-DEC-1999.

XX 21-MAY-1999; 99WO-AU00385.

XX 21-MAY-1998; 98AU-0003634.

XX (UNSY) UNIV SYDNEY.

XX Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX Novel nucleic acid molecule useful for the detection of flagellated

XX bacterial strains in food, faeces, etc.

XX Claim 3; Page 225; 245pp; English.

XX AAZ56331 to AAZ56398 represent nucleic acid molecules (1) encoding all
or part of an Escherichia coli flagellin protein except a protein
expressed by E. coli H1, H7, H12 or H48 type strains. The present
invention also describes a method of detecting the presence of E. coli
of a particular H serotype in a sample, comprising specifically
hybridising a nucleic acid, preferably at least a pair, derived from a
flagellating gene, specific for a particular flagellin gene associated
with the H serotype, to any E. coli in the sample which contain the gene,
and detecting any hybridised molecules, identifying the presence of that
serotype in the sample. (1) are useful for: (1) detecting the presence
of E. coli of H serotype in a sample by hybridising at least one or a
pair of (1) to any E. coli in the sample and detecting the hybridised
nucleic acid molecules; and (2) for detecting the presence of both O
and H-serotypes of E. coli by hybridising at least one or a pair of (1)
to any E. coli present in the sample and detecting the hybridised
nucleic acid molecules. (1) is particularly useful for detecting the
combination of O and H antigen. Hybridised (1) when using at least one
(1) is detected by southern blot analysis and, when using a pair of (1),
is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
represent primers used in the exemplification of the present invention.

XX Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 1344;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcatcgccagcatt 20

Db 590 GCATCGCCAGCATT 577

RESULT 14

AAA69762/c

ID AAA69762 standard; cDNA; 2017 BP.

XX AAA69762;

XX 07-NOV-2000 (first entry)

XX Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

XX Human ovarian carcinoma; ovarian cancer; therapy; diagnosis;

XX Tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.

XX Homo sapiens.

XX WO200036107-A2.

XX 22-JUN-2000.

XX 17-DEC-1999; 99WO-US30270.

XX 17-DEC-1998; 98US-0215681.

XX 17-DEC-1998; 98US-0216003.

XX 23-JUN-1999; 99US-0338933.

XX 24-SEP-1999; 99US-0404879.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, King GE, Algate PA, Frudakis TN;

XX WPI; 2000-431589/37.

XX Immunogenic portion of an ovarian carcinoma protein and the nucleic

XX acid encoding it, useful for the diagnosis, prevention and treatment of

XX cancer, preferably ovarian cancer

XX Claim 1; Fig 2; 299pp; English.

XX The present invention describes an isolated polypeptide comprising an

XX immunogenic portion of an ovarian carcinoma protein (or its variants).

XX Ovarian carcinoma proteins, and polynucleotides encoding them, have

XX cytostatic activity and can be used in gene therapy and vaccines.

XX Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines

XX are useful for the prevention, diagnosis and treatment of cancer,

XX preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557

XX represent human ovarian carcinoma polynucleotides and proteins used in

XX the exemplification of the present invention.

XX Sequence 2017 BP; 673 A; 440 C; 462 G; 442 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 2017;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 agtgcacgcagc 16

Db 964 AGTGCACGCAGC 951

RESULT 15

AAT30684/c

ID AAT30684 standard; DNA; 2193 BP.

XX AAT30684;

XX 11-SEP-1996 (first entry)

XX Kaposi's sarcoma associated herpesvirus ORF22.

XX Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;

XX therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

XX kaposi's sarcoma associated herpesvirus.

XX WO9615779-A1.

XX 30-MAY-1996.

XX 21-NOV-1995; 95WO-US15138.
 XX 11-APR-1995; 95US-0420235.
 PR 21-NOV-1994; 94US-0343101.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA Chang Y, Moore PS;
 PI WPI: 1996-268320/27.
 DR P-PSDB; AAR97832.
 PT Herpes virus associated with Kaposi's sarcoma - also definitive DNA
 PT sequences, useful for diagnosis of and to develop prods. for
 PT treatment of Kaposi's sarcoma
 XX
 PS Claim 17; Page 198-201; 277pp; English.
 XX
 CC Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),
 CC 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2093-
 CC 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment
 CC of a newly identified human gamma-2 herpesvirus associated with
 CC Kaposi's sarcoma (KS). The ORFs were identified by sequence and
 CC positional homology to known herpesvirus sequences. Probes that bind
 CC specifically to the KS associated herpesvirus can be used for KS
 CC diagnosis. Antisense or triplex oligonucleotides are useful for
 CC prophylaxis or treatment of KS, and the protein products (see also
 CC AAR97830-46) of the 17 open reading frames are useful as vaccines.
 XX
 SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
 ||||||||||||
 DB 46 TGCATGCCACGACAT 33

RESULT 16
 AAT16814/C
 ID AAT16814 standard; DNA: 2193 BP.
 XX
 AC AAT16814;
 XX
 DT 13-AUG-1996 (first entry)
 XX
 DE Kaposi's sarcoma associated herpesvirus ORF22.
 XX
 KW Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
 KW glycoprotein H; ss.
 XX
 OS Kaposi's sarcoma associated herpesvirus.
 XX
 PN WO9606159-A1.
 XX
 PD 29-FEB-1996.
 XX
 PF 11-AUG-1995; 95WO-US10194.
 XX
 PR 11-APR-1995; 95US-0420235.
 PR 18-AUG-1994; 94US-0292365.
 PR 21-NOV-1994; 94US-0343100.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRANT) GRANT D E.
 PA (VIELE) VIELE L.
 XX
 PI Chang Y, Moore PS, Grant DE, Viele L;
 XX

DR WPI: 1996-151362/15.
 DR P-PSDB; AAR93608.
 XX
 PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated
 PT vectors and proteins, used in detection and vaccination.
 XX
 PS Claim 17; Page 211-214; 305pp; English.
 XX
 CC Kaposi's sarcoma associated herpes virus (KSHV) clone KS5 (AAT16806),
 CC obld. from a KS lesion genomic library, includes 15 complete ORFs and
 CC 2 incomplete ORFs (AAT16807-23) named according to their herpesvirus
 CC salmuri positional homologues. ORF22 (AAT16814) codes for
 CC glycoprotein H (AAR93608). KSHV DNA may be incorporated into a vector
 CC and expressed in host cells to produce peptides useful in vaccines or
 CC for raising antibodies. The DNA may itself be used to immunise a
 CC subject, or used to design therapeutic antisense and triplex
 CC molecules or diagnostic probes, or to raise transgenic animals.
 XX
 SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
 ||||||||||||
 DB 46 TGCATGCCACGACAT 33

RESULT 17
 AAT20914/C
 ID AAT20914 standard; DNA: 2193 BP.
 XX
 AC AAT20914;
 XX
 DT 02-DEC-1999 (first entry)
 XX
 DE Nucleotide sequence of gH glycoprotein of HHV-8.
 XX
 KW vaccine; antibody; Kaposi's sarcoma; glycoprotein; detection;
 KW HHV-8; Human Herpesvirus type 8; ds.
 XX
 OS Human Herpes Virus type 8.
 XX
 PN WO9445948-A1.
 XX
 PD 16-SEP-1999.
 XX
 PF 12-MAR-1999; 99WO-US05464.
 XX
 PR 12-MAR-1998; 98US-0077670.
 XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA
 PI Whitbeck JC, Eisenberg RJ, Cohen GH;
 XX
 DR WPI: 1999-551211/46.
 DR P-PSDB; AAY42395.
 XX
 PT New nucleic acid molecules encoding human herpesvirus type glycoprotein
 PT H and glycoprotein L polypeptides for diagnosis and treatment of
 PT Kaposi's Sarcoma -
 XX
 PS Claim 3; Fig 1; 75pp; English.
 XX
 CC This is the nucleotide sequence of the human herpes virus type 8
 CC glycoprotein H. This sequence and it's encoded protein (AAY42395) can be
 CC used in the development of vaccines, antibodies and other preparations.
 CC The vaccine is useful as a therapeutic agent for prevention of HHV-8
 CC disease and as a therapeutic agent for treatment of HHV-8 infection
 CC in order to boost the immune response in the animal. Administration of
 CC a pure preparation of soluble HHV-8 gH/gL complex or at least one

CC polynucleotide encoding the soluble complex is useful for preventing or
 CC treating an HIV-8 infection in a human.
 CC Diagnosing of an HIV-8 infection in a human is carried out by contacting
 CC a biological sample with an antibody which specifically binds to either
 CC glycoprotein or in complex form. If binding of the antibody occurs then
 CC the human has HIV-8 infection.
 CC Alternatively the infection is diagnosed by addition of one of the
 CC HIV-8 glycoproteins or complex and determining whether specific binding
 CC of the protein to an antibody in the biological sample occurs. A further
 CC diagnostic method involves contacting the biological sample with an
 CC HIV-8 glycoprotein H polynucleotide or HIV-8 glycoprotein L
 CC polynucleotide and determining whether binding to a nucleic acid in the
 CC sample occurs.
 XX
 SO Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcgcagcat 19
 |||||
 Db 46 TGCATGCCGACGACAT 33

RESULT 18

AAZ39011/C
 ID AAZ39011 standard; cDNA; 3593 BP.

AC AAZ39011;

DT 28-FEB-2000 (first entry)

DE Mouse Ese2 coding sequence.

KW Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;
 regulation; actin cytoskeleton; detection; cancer; infection;
 EH-domain and SH3-domain regulator of endocytosis; anticancer;
 antiproliferative; antiviral; ss.

OS Mus sp.

PN W09955728-A2.

PD 04-NOV-1999.

PF 27-APR-1999; 99WO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

PA (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A;

DR WPI: 2000-052802/04.

DR P-PSDB: AAY57445.

PT New nucleic acid encoding Esel and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 infection

XX Claim 25; Page 46-48; 99pp; English.

XX The present sequence encodes mouse Ese2. The present invention
 CC specifically describes mammalian Esel and 2 proteins (I) and their splice
 CC variants (Ese - EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or

CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Esel is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 SO Sequence 3593 BP; 1171 A; 783 C; 850 G; 789 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agtgcacgcgcag 16
 |||||
 Db 923 AGTGCATGCCGACG 910

RESULT 19

AAZ49334
 ID AAZ49334 standard; cDNA; 4189 BP.

AC AAZ49334;

DT 14-MAR-2000 (first entry)

DE Murine multidrug resistance-1 (MDR-1) cDNA.

KW Multidrug resistance; MDR-1; P-glycoprotein;
 transmembrane efflux pump; hematopoietic stem cell; transduction;
 bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 gene therapy; gene replacement; genetic defect; thalassemia;
 Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 cytokine; ds.

OS Mus sp.

PN W09961589-A2.

PD 02-DEC-1999.

PF 27-MAY-1999; 99WO-US11825.

PR 28-MAY-1998; 98US-0086988.

PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Sorrentino B, Bunting K;

DR WPI: 2000-072615/06.

DR P-PSDB: AAY58188.

PT Ex vivo expansion of hematopoietic stem cells transduced with a
 sequence encoding human multidrug resistance-1, used for bone marrow
 transplantation

XX Disclosure; Page 90-93; 113pp; English.

XX This sequence represents cDNA encoding murine multidrug
 CC resistance protein MDR-1. MDR-1 is a transmembrane

CC efflux pump, responsible for the export of drugs from cells,
 CC particularly cancer cells. The invention relates to transducing
 CC hematopoietic stem cells with nucleic acid encoding an MDR protein
 CC and culturing the modified cells. The modified hematopoietic stem
 CC cells are useful in bone marrow transplantation (to reconstitute
 CC hematopoietic systems in patients who have undergone chemotherapy or
 CC radiation therapy) and in ex vivo gene therapy of genetic defects in
 CC cells derived from hematopoietic stem cells, e.g., thalassemia,
 CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
 CC cells can also be used to identify factors involved in regulating
 CC proliferation and differentiation in hematopoietic stem cells.
 CC Hematopoietic stem cells that express MDR-1 will be protected against
 CC chemotherapeutic agents, so can be engrafted while the patient is
 CC undergoing chemotherapy. Expansion of (rare) hematopoietic stem cells
 CC provides sufficient cells to permit standard biochemical analysis.
 CC Overexpression of MDR-1 allows cytokine-driven expansion of
 CC hematopoietic stem cells by at least 10-fold compared with a maximum
 CC of 4-fold in known procedures.

XX
 SO Sequence 4189 BP; 1204 A; 875 C; 1028 G; 1082 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4189;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagcattcgat 25
 |||||
 Db 880 gccagcattcgat 893

RESULT 20
 AAQ38950
 ID AAQ38950 standard; DNA; 4313 BP.
 AC AAQ38950;
 XX
 DT 28-JUL-1993 (first entry)
 DE Mouse multidrug resistance sequence.
 XX
 KW mdr gene; lambda DR11 clone; ss.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 110..3940
 FT /*tag= a
 FT /phenotype= multidrug_resistance
 XX
 FN US5198344-A.
 XX
 PD 30-MAR-1993.
 XX
 PF 15-JUL-1986; 86US-0885951.
 XX
 PR 15-JUL-1986; 86US-0885951.
 PR 06-FEB-1991; 91US-0652311.
 XX
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 PI Croop JM, Gros P, Housman DE;
 XX
 DR WPI; 1993-126077/15.
 DR P-PSDB; AAR35199.
 XX
 PT DNA sequence which confers multi-drug resistance on sensitive
 PT mammalian cells - used to preserve bone marrow cells during
 PT chemotherapy to prevent infection
 XX
 PS Claim 1; Fig 8; 22pp; English.
 XX
 CC A cDNA library was constructed from mRNA which had been isolated

CC from a drug-sensitive mouse cell-line. Two mouse cDNA molecules
 CC complementary to the mRNA species encoded by 2 related but distinct
 CC cDNAs were isolated and cloned. One of the cDNA clones (lambda
 CC DR11) is a full-length cDNA clone for one member of the mdr gene
 CC family. When incorporated into prokaryotic expression vector pUREX4
 CC (which allows high levels of transcription of the cDNA when
 CC introduced into mammalian cells), the clone was shown to confer the
 CC multidrug resistance phenotype upon transfection into drug-sensitive
 CC mammalian cells.

XX
 SO Sequence 4313 BP; 1241 A; 904 C; 1058 G; 1110 T; 0 other;

Query Match 56.0%; Score 14; DB 14; Length 4313;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagcattcgat 25
 |||||
 Db 989 gccagcattcgat 1002

RESULT 21
 AAZ39010/C
 ID AAZ39010 standard; cDNA; 4625 BP.
 AC AAZ39010;
 XX
 DT 28-FEB-2000 (first entry)
 DE Mouse Ese2 full length cDNA sequence.
 XX
 KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 XX
 PN W09955728-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;
 XX
 DR WPI; 2000-052802/04.
 DR P-PSDB; AAY57445.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 25; Page 43-46; 99pp; English.
 XX
 CC The present sequence encodes mouse Ese2. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while

CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC XX

Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4625;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcag 16
 |||
 Db 1173 AGGTGCATCGCCAG 1160

RESULT 22
 AAZ39027/C
 ID AAZ39027 standard; CDNA; 4975 BP.
 XX
 AC AAZ39027;
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Ese2L coding sequence.
 XX
 KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 XX
 PN WO995728-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;
 XX
 DR WPI: 2000-052802/04.
 DR P-PSDB; AAY57450.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 25; Page 66-69; 99pp; English.
 XX

CC The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antinodules (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of

CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese2L coding sequence.
 CC XX

Sequence 4975 BP; 1540 A; 1166 C; 1201 G; 1068 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4975;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcag 16
 |||
 Db 923 AGGTGCATCGCCAG 910

RESULT 23
 AAV35235/C
 ID AAV35235 standard; DNA; 5319 BP.
 XX
 AC AAV35235;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE R. prowazekii S-layer protein genomic DNA.
 XX
 KW Surface layer protein; S-layer; vaccine; typhus; spotted fever;
 KW infection; diagnosis; disease; ds.
 XX
 OS Rickettsia prowazekii.
 XX
 FH key Location/Qualifiers
 FT -35_signal 340..345
 FT /*tag= a
 FT 363..368
 FT /*tag= b
 FT 379..386
 FT /*tag= c
 FT 391..5229
 FT /*tag= d
 FT /*product= S-layer protein
 FT 5270..5306
 FT /*tag= e
 XX
 PN US5783441-A.
 XX
 PD 21-JUL-1998.
 XX
 PF 20-DEC-1993; 93US-0169927.
 XX
 PR 20-DEC-1993; 93US-0169927.
 PR 09-AUG-1991; 91US-0742128.
 XX
 PA (USNA) US SEC OF NAVY.
 XX
 PI Carl M, Ching W, Dasch GA, Dobson ME;
 XX
 DR WPI: 1998-427031/36.
 DR P-PSDB; AAW65088.
 XX
 PT Recombinant DNA encoding Rickettsia surface layer proteins - useful
 PT for diagnosing typhus and spotted fever and for preparing vaccines
 PT against them
 XX
 PS Claim 1; Column 11-24; 20pp; English.
 XX
 CC This sequence encodes the Surface layer (S-layer) protein from

CC R. prowazekii strain Breinl. This sequence is useful for vaccination
CC against typhus and spotted fever rickettsial infection or for diagnosing
CC diseases caused by these bacteria. The surface layer protein antigens can
CC be produced recombinantly in large quantities.

XX
SQ Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5319;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 catgccagcatt 21
|||||
Db 2221 CATGCCAGCATTT 2208

RESULT 24
AAV21186/c
ID AAV21186 standard; DNA: 5676 BP.

XX AAV21186;

XX 24-JUL-1998 (first entry)

XX Amycolatopsis mediterranei strain W3136 5.7 kb KpnI DNA fragment.

XX Amycolatopsis mediterranei: rifamycin; synthesis: gene cluster;

KW polypeptide synthase; actinomycete; ansamycin; ds.

XX Amycolatopsis mediterranei.

XX Key Location/Qualifiers

XX CDS

XX 3..5676
XX /*tag= a
XX /product= "polypeptide synthase"
XX /note= "no stop codon given"

XX W09807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97MO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schnupp T, Toupet C;

XX WPI: 1998-169172/15.

XX P-PSDB: AAW52844.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues

XX Claim 4; Page 38-43; 205pp; English.

XX The present sequence represents Amycolatopsis mediterranei strain W3136
XX 5.7 kb KpnI fragment DNA, from the present invention. The present
XX invention describes a Amycolatopsis mediterranei rifamycin synthesis
XX gene cluster DNA fragment comprising a DNA region involved directly or
XX indirectly in the gene cluster responsible for rifamycin synthesis,
XX including the adjacent DNA regions to the right and left which, by
XX reason of their function in connection with rifamycin biosynthesis,
XX qualify as constituents of this rifamycin gene cluster, and functional
XX fragments, derivatives or constituents of these. The Amycolatopsis
XX mediterranei rifamycin synthesis gene cluster DNA fragment can be used
XX for producing rifamycin, rifamycin analogues or precursors. It can also
XX be used for inactivating or modifying genes involved in ansamycin or
XX rifamycin biosynthesis. The DNA can be used for constructing mutant
XX actinomycetes strains from which the natural rifamycin or ansamycin
XX biosynthesis gene cluster has been partly or completely deleted. The

CC DNA fragment can be used for assembling a library of polypeptide
CC synthases, which can be used for assembling a library of polypeptides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX
SQ Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5676;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caagtgatcgcca 15
|||||
Db 3923 CAGTGCATCGCCA 3910

RESULT 25
AAZ39026/c
ID AAZ39026 standard; cDNA: 6014 BP.

XX AAZ39026;

XX 28-FEB-2000 (first entry)

XX Mouse Ese2L cDNA sequence.

XX Mouse; murine; Ese2; endocytosis; vesicular trafficking;

KW regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;

XX antiproliferative; antiviral; ss.

XX Mus sp.

XX W09955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99MO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 9905-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI: 2000-052802/04.

XX P-PSDB: AAY57450.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection

XX Claim 25; Page 63-66; 99pp; English.

XX The present invention specifically describes mammalian Ese1 and 2
XX proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
XX regulator of endocytosis). (I) are involved in regulation of clathrin-
XX mediated endocytosis (as a complex with Eps15 protein), vesicular
XX trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
XX mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
XX sequences antisense to the (I) polynucleotide; agents that downregulate
XX expression of Ese genes or antagonists of an Ese binding partner are
XX used to treat diseases associated with undesirable endocytosis and
XX resulting changes in cellular function. Particularly overexpression of
XX Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
XX cultures, while administration of (I) is used to promote endocytosis of
XX selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
XX proliferation of cells that can be stimulated to proliferate by a growth
XX factor receptor, and similar compounds (also inactive Ese mutants) can be
XX used to prevent vital infection. Endocytosis may also be regulated, in
XX vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding

DT 13-AUG-1996 (first entry)
 XX Kaposi's sarcoma associated herpesvirus clone KS5.
 DE Kaposi's sarcoma; KSHV; Lymphoma; AIDS; vaccine; diagnosis; therapy;
 XX ss.
 KW Kaposi's sarcoma associated herpesvirus.
 OS WO9606159-A1.
 PN 29-FEB-1996.
 PD 11-AUG-1995; 95WO-0510194.
 PF 11-APR-1995; 95US-0420235.
 PR 18-AUG-1994; 94US-0292365.
 PR 21-NOV-1994; 94US-0343100.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRANT/) GRANT D E.
 PA (VIEL/) VIELE L.
 XX Chang Y, Moore PS, Grant DE, Viele L;
 PI WPI: 1996-151362/15.
 DR Herpesvirus DNA associated with Kaposi's sarcoma - also associated
 XX vectors and proteins, used in detection and vaccination.
 PT Claim 16: Page 177-188; 305pp; English.
 PS Representational difference analysis was used to identify unique
 XX DNA sequences in Kaposi's sarcoma (KS) tissue. An isolated sequence,
 CC KS3308am (AA116824), was used to screen a KS lesion genomic library. A
 CC lambda phage, KS5, was identified and the sequence of its insert
 CC (AA116808) was detd. This represents a novel infectious agent
 CC associated with AIDS-KS, Kaposi's sarcoma associated herpesvirus (KSHV).
 CC The KS5 fragment has 15 complete ORFs and 2 incomplete ORFs (see also
 CC AA116807-23). Portions of the sequence may be incorporated into a
 CC vector and expressed in host cells to produce peptides (see also
 CC AA93601-17) useful in vaccines or for raising antibodies. The DNA may
 CC itself be used to immunise a subject, or used to design therapeutic
 CC antisense and triplex molecules or diagnostic probes, or to raise
 CC transgenic animals.
 XX SQ Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 20710;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacagcat 19
 ||||||||||||
 DB 2138 TGCATCGCCAGCAT 2125

RESULT 28
 AAV73803/C
 ID AAV73803 standard; DNA: 35100 BP.
 XX AAV73803;
 AC 25-FEB-1999 (first entry)
 DE KSHV LUR DNA (nucleotides 35,101-70,200).
 XX Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
 KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
 KW diagnosis; treatment; HHV8; thymidine kinase; glycoprotein M;
 KW capsid protein; packaging protein; helicase primase; glycoprotein M;
 KW viral protein kinase; alkaline exonuclease; virion assembly protein;
 XX FH

KW uracil DNA glycosylase; UDG; glycoprotein L; ds.
 XX Kaposi's sarcoma-associated herpesvirus.
 OS US5849564-A.
 XX 15-DEC-1998.
 PN 29-NOV-1996; 96US-0770379.
 PD 29-NOV-1996; 96US-0770379.
 PF 29-NOV-1996; 96US-0770379.
 PR (UYCO) UNIV COLUMBIA NEW YORK.
 PA Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
 PI WPI: 1999-069741/06.
 DR Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
 PT dihydrofolate reductase and is useful for treatment, prophylaxis
 PT or diagnosis of Kaposi's sarcoma
 XX Disclosure; Column 97-126; 109pp; English.
 XX This sequence is a fragment of the Kaposi's sarcoma-associated
 CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
 CC coding regions for ORF20, ORF21 which encodes thymidine kinase, ORF22
 CC which encodes glycoprotein H, ORF23, ORF24, ORF25 which encodes the
 CC major capsid protein, ORF26 which encodes capsid protein II, ORF27,
 CC ORF28, ORF29b which encodes packaging protein II, ORF30, ORF31, ORF32,
 CC ORF33, ORF29a which encodes packaging protein I, ORF34, ORF35, ORF36
 CC which encodes viral protein kinase, ORF37 which encodes alkaline
 CC exonuclease, ORF38, ORF39 which encodes glycoprotein M, ORF40 which
 CC encodes helicase primase subunit I, ORF41 which encodes helicase primase
 CC subunit II, ORF42, ORF43 which encodes capsid protein III, ORF44 which
 CC encodes helicase primase subunit III, ORF45 which encodes the virion
 CC assembly protein, ORF46 which encodes uracil DNA glycosylase (UDG),
 CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
 CC form of neoplasm occurring in persons with acquired immune deficiency
 CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
 CC diagnosis and treatment of a subject with Kaposi's sarcoma and for
 CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a
 CC cell.
 XX SQ Sequence 35100 BP; 8016 A; 9231 C; 9477 G; 8376 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 35100;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacagcat 19
 ||||||||||||
 DB 2058 TGCATCGCCAGCAT 2045

RESULT 29
 AAV21187/C
 ID AAV21187 standard; DNA: 53789 BP.
 XX AAV21187;
 AC 24-JUL-1998 (first entry)
 DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
 XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 KW polyketide synthase; actinomycete; ansamycin; ds.
 XX Amycolatopsis mediterranei.
 OS Key Location/Qualifiers
 XX FH

PT the presence of a sequence polymorphism -
 XX
 PS Claim 1; Page 269; 543pp; English.
 XX
 CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
 CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
 CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
 CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
 CC consecutive pairs of nucleotides containing SNPs which result in changes
 CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
 CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
 CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
 CC in non-conservative changes. The SNPs in sequences 1187 to 1192
 CC (AAA77504-A77509) generate frameshift mutations. The invention also
 CC relates to a method of detecting a polymorphic site in a nucleic acid and
 CC a method of determining the relatedness of two nucleic acids. It also
 CC encompasses peptides containing polymorphic sites, antibodies raised
 CC against such peptides, and a method of detecting polymorphic
 CC proteins/peptides using the antibodies. The nucleic acids are useful for
 CC gene therapy of an individual having, suspected of having, or at risk of
 CC developing a pathological condition due to the presence of a sequence
 CC polymorphism. Such treatment would comprise administration of the
 CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
 CC peptides can also be used in the treatment of such individuals.
 XX
 SQ Sequence 51 BP; 13 A; 11 C; 15 G; 12 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 cgcacgacattcg 23
 |||||
 DB 11 cgcacgacattcg 23

RESULT 33
 AAA76687
 ID AAA76687 standard; cDNA; 51 BP.
 XX
 AC AAA76687;
 XX
 DT 16-NOV-2000 (first entry)
 XX
 DE Human clone cg28389525 polymorphic site, SEQ ID NO:370.
 XX
 KW Human: single nucleotide polymorphism; SNP;
 KW detection; identification; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT replace (26,T)
 FT variation /*tag= a
 XX
 PN WO200029623-A2.
 XX
 PD 25-MAY-2000.
 XX
 PF 17-NOV-1999; 99WO-US27293.
 XX
 PR 17-NOV-1998; 98US-0109024.
 PR 16-NOV-1999; 99US-0109024.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkels RA, Leach MD;
 XX
 DR WPI: 2000-387826/33.
 XX
 PT Human nucleic acids containing single nucleotide polymorphisms, useful
 PT for treating a subject suffering, or at risk from a pathology due to

PT the presence of a sequence polymorphism -
 XX
 PS Claim 1; Page 270; 543pp; English.
 XX
 CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
 CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
 CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
 CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
 CC consecutive pairs of nucleotides containing SNPs which result in changes
 CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
 CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
 CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
 CC in non-conservative changes. The SNPs in sequences 1187 to 1192
 CC (AAA77504-A77509) generate frameshift mutations. The invention also
 CC relates to a method of detecting a polymorphic site in a nucleic acid and
 CC a method of determining the relatedness of two nucleic acids. It also
 CC encompasses peptides containing polymorphic sites, antibodies raised
 CC against such peptides, and a method of detecting polymorphic
 CC proteins/peptides using the antibodies. The nucleic acids are useful for
 CC gene therapy of an individual having, suspected of having, or at risk of
 CC developing a pathological condition due to the presence of a sequence
 CC polymorphism. Such treatment would comprise administration of the
 CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
 CC peptides can also be used in the treatment of such individuals.
 XX
 SQ Sequence 51 BP; 13 A; 12 C; 15 G; 11 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 cgcacgacattcg 23
 |||||
 DB 11 cgcacgacattcg 23

RESULT 34
 AAA45502/c
 ID AAA45502 standard; cDNA; 286 BP.
 XX
 AC AAA45502;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:2077.
 XX
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
 KW cerebroprotective; neuroprotective; nootropic; antiproliferative;
 KW vaccine; autoimmune disorder; antidepressant; gene therapy;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX
 PR 15-OCT-1998; 98US-0104436.
 XX
 PA (GENY) GENETICS INST INC.

XX
PI Jacobs K, McCoy JM, Lavallee ER, Collins-Racie IA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
XX
DR WPI: 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1, Page 723; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cyostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiallergic; antitumor; osteopathic; neuroprotective;
CC neurotrophic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 286 BP; 44 A; 84 C; 69 G; 89 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgc 13
|||
Db 93 GCAGGTGATCC 81

RESULT 35
AAZ14744/C
ID AAZ14744 standard; cDNA: 300 BP.
XX
AC AAZ14744;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:2213.
XX
XX Human: gene: gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX
PN W09938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99MO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
PI Jones WJ, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Grain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
XX
PS Claim 1, Page 1159; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcacgcacgcat 19
|||
Db 174 GCATGCCACGAT 162

RESULT 36
AAZ14726
ID AAZ14726 standard; cDNA: 300 BP.
XX
AC AAZ14726;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:2195.
XX
XX Human: gene: gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX
PN W09938972-A2.
XX
PD 05-AUG-1999.
XX

PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080315.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Suduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1; Page 1155; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA12532 to AA17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA12532 to AA17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 CC
 SQ Sequence 300 BP; 80 A; 64 C; 63 G; 93 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcagtgcatcgc 13
 |||||
 DB 180 gcagtgcatcgc 192

RESULT 37
 AAA31174/C
 ID AAA31174 standard; DNA; 322 BP.
 XX
 AC AAA31174;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #135.
 XX
 KW Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Euca1yptus grandis.

XX
 PN WO967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 25-JUN-1999; 99WO-N200092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukkala JJ, Bloksberg LN, Glenn M;
 PI WPI; 2000-116958/10.
 XX
 DR
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX
 PS Claim 1; Page 116; 392pp; English.
 XX
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 CC
 SQ Sequence 322 BP; 43 A; 101 C; 88 G; 89 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atcgccagcatTT 21
 |||||
 DB 21 ATCGCCAGCATTT 9

RESULT 38
 AAT27962
 ID AAT27962 standard; DNA; 447 BP.
 XX
 AC AAT27962;
 XX
 DT 11-MAR-1997 (first entry)
 XX
 DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
 XX
 KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection; ss.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9613590-A2.
 XX
 PD 09-MAY-1996.
 XX
 PF 23-OCT-1995; 95WO-EP04155.
 XX
 PR 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Stuyver L;

XX WPI: 1996-251460/25.
 DR P-PSDB: AAR96551.
 XX
 PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
 PT - used to develop probes and primers for new sub:types and vaccines
 PT to prevent and treat infection
 XX
 PS Claim 6; Fig 3; 150pp; English.
 XX
 CC The sequences AAT27937-T72989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the HCV
 CC genome. This sequence represents nucleotides 478-925 from the HCV type
 CC 10a isolate NE98.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 XX
 SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 3 other;

Query Match 52.0%; Score 13; DB 17; Length 447;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccgcagcatt 20
 |||||
 DB 406 catgccgcagcatt 418

RESULT 39
 AAZ40788
 ID AAZ40788 standard; DNA: 538 BP.
 XX
 AC AAZ40788;

DT 18-JAN-2000 (first entry)
 XX

DE Secreted protein EST coding sequence 108-003-5-0-H2-FL.

XX Secreted protein; fingerprint identification technique;
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.

XX Homo sapiens.

OS WO9940189-A2.

PN 12-AUG-1999.

PD 09-FEB-1999; 99WO-1B00282.

PF 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PR 04-SEP-1998; 98US-0099273.

XX (GEST) GENSET.
 XX Bougueleret L, Duclert A, Dumas Mline Edwards J;
 XX

XX WPI: 1999-600966/51.
 DR P-PSDB: AAF59660.
 XX

PT Extended cDNAs useful for expressing secreted proteins and to obtain
 PT specific antibodies -
 XX

PS Claim 1; Page 154-155; 244pp; English.

XX This sequence encodes a human secreted protein of the invention. The
 CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
 CC prepare PCR primers and probes. These are useful for forensic matching or
 CC positive identification by DNA sequencing. They may also be used in
 CC alternative fingerprint identification techniques. Antibodies against the
 CC proteins encoded by the extended cDNAs are useful in identification of
 CC tissue types or cell species, as well as identifying tissue specific
 CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
 XX
 SQ Sequence 538 BP; 104 A; 159 C; 151 G; 124 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 538;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgc 13
 |||||
 DB 314 gcaggtgcatgc 326

RESULT 40
 AAF09259
 ID AAF09259 standard; cDNA: 629 BP.
 XX
 AC AAF09259;

DT 13-MAR-2001 (first entry)
 XX

DE Fusarium venenatum EST SEQ ID NO:1782.

XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Fusarium venenatum.

OS WO200056762-A2.

PN 28-SEP-2000.

PD 22-MAR-2000; 2000WO-US07781.

PF 22-MAR-1999; 99US-0273623.

PR (NOVO) NOVO NORDISK INC.

PR (NOVO) NOVO NORDISK AS.

PR Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 XX
 PS Claim 86; Page 1051; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC niger; AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SO Sequence 629 BP; 153 A; 169 C; 135 G; 168 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 629;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catcgccagcatt 20
 |||||
 Db 83 catcgccagcatt 95

RESULT 41
 AAF09268/C
 ID AAF09268 standard; cDNA; 632 BP.
 XX
 AC AAF09268;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:1791.
 XX
 KM Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX
 DR WPI; 2000-594572/56.
 XX

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 XX
 PS Claim 86; Page 1054; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC niger; AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SO Sequence 632 BP; 168 A; 168 C; 136 G; 148 T; 12 other;

Query Match 52.0%; Score 13; DB 21; Length 632;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 ggtgcacgcagc 16
 |||||
 Db 547 GGTGCATCGCCAG 535

RESULT 42
 AAC55942/C
 ID AAC55942 standard; DNA; 646 BP.
 XX
 AC AAC55942;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Eucalyptus grandis transcription factor DNA sequence #73.
 XX
 KM Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KM poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KM homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
 KM type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI: 2000-579369/54.
 DR
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 XX
 PS Claim 1, Page 63; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeo/c/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SO Sequence 646 BP; 129 A; 195 C; 174 G; 148 T; 0 other;
 Query Match 52.0%; Score 13; DB 21; Length 646;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcacgc 17
 |||||||||
 DB 496 CTCGATCCCGCAGC 484

RESULT 43
 AA217005
 ID AA217005 standard; cDNA: 714 BP.
 XX
 AC AA217005;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:4475.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Gliese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 XX WPI: 1999-494092/41.
 DR
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PT
 XX
 PS Claim 1, Page 2120; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX
 SO Sequence 714 BP; 186 A; 142 C; 146 G; 231 T; 9 other;

Query Match 52.0%; Score 13; DB 20; Length 714;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcgccagatttc 22
 |||||||||
 DB 74 tcgccagatttc 86

RESULT 44
 AAF22436
 ID AAF22436 standard; cDNA: 725 BP.
 XX
 AC AAF22436;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:15.
 XX
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200073801-A2.
 XX
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000WO-US14749.
 XX
 PR 28-MAY-1999; 99US-0136526.
 PR 10-SEP-1999; 99US-0133454.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Obata Y;
 XX
 DR WPI: 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition
 PT characterized by expression of an abnormal amount of a protein, e.g.
 cancer -
 PS Claim 50; Page 276; 799pp; English.
 XX
 CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
 CC represent nucleotide sequences encoding human breast, gastric and
 CC prostate cancer associated antigen precursors (CAAP) respectively.
 CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
 CC represent human breast, gastric and prostate CAAP protein sequence
 CC respectively. CAAPs have cytostatic activity and can be used in the
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
 CC condition characterised by expression of an abnormal amount of a protein,
 CC e.g. cancer.
 CC
 SQ Sequence 725 BP; 131 A; 225 C; 224 G; 136 T; 9 other;
 Query Match 52.0%; Score 13; DB 22; Length 725;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatgccagcat 19
 |||||||||
 Db 229 gcatgccagcat 241
 RESULT 45
 AA216470
 ID AA216470 standard; cDNA; 749 BP.
 XX
 AC AA216470;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:3940.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9338972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99MO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 PA
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX

PS Claim 1; Page 1872; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 CC
 SQ Sequence 749 BP; 182 A; 138 C; 154 G; 258 T; 17 other;
 Query Match 52.0%; Score 13; DB 20; Length 749;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gccagtgcatgc 13
 |||||||||
 Db 260 gccagtgcatgc 272

Search completed: October 9, 2001, 15:52:31
 Job time: 15112 sec

Wed Oct 10 07:45:58 2001

us-09-396-196f-6.011.rng

Page 27

Query Match 100.0%; Score 25; DB 2; Length 1041;

Best Local Similarity 100.0%; Pred. No. 6.5e-06;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgccagcatttcgat 25
|||||
DB 84 GCAGGTGTCATCGCCAGCATTTCGAT 108

RESULT 2

US-08-846-338-7

Sequence 7, Application US/08846338

Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 5869719artis Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8589

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;

Best Local Similarity 100.0%; Pred. No. 6.5e-06;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgccagcatttcgat 25
|||||
DB 84 GCAGGTGTCATCGCCAGCATTTCGAT 108

RESULT 3

US-08-411-768B-1

Sequence 1, Application US/08411768B

Patent No. 6083712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw

TITLE OF INVENTION: Biotechnological Method

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect

VERSION: 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/411,768B

FILING DATE: 31-March-95

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 3124/92

FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 2134/93

FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5872 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

STRAIN: DSM498

IMMEDIATE SOURCE:

CLONE: pBO30A-15/9

FEATURE:

NAME/KEY: CDS

LOCATION: 117..1157

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 117

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioB"

OTHER INFORMATION: /number= 1

FEATURE:

NAME/KEY: CDS

LOCATION: 2295..3050

OTHER INFORMATION: /codon_start= 2295

OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"

OTHER INFORMATION: /product= "protein"

OTHER INFORMATION: /gene= "bioC"

OTHER INFORMATION: /number= 3

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: aminotransf."

FEATURE:

NAME/KEY: CDS


```
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no. 9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match      100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 gcaggtgcatgcgcagcattcgat 25
        |||
Db      200 GCAGGTGCATCGCCAGCATTTCGAT 224

RESULT      4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
```

```
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSER: Brunhaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYDROTHERMAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
PUBLICATION NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
 PUBLICATION DATE: 07-APR-1993
 US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 6,7e-06;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacgcatcgcacat 25
 ||||||||||||||||||
 DB 200 GCAGGTGCATGCCAGCATTCGAT 224

RESULT 5
 US-09-142-078-45/c
 ; Sequence 45, Application US/09142078
 ; Patent No. 6172041
 ; GENERAL INFORMATION:
 ; APPLICANT: McCabe, R. Tyler
 ; APPLICANT: Zhou, Li-Ming
 ; APPLICANT: Layer, Richard T.
 ; TITLE OF INVENTION: Use of Conantokins
 ; NUMBER OF SEQUENCES: 71
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.
 ; STREET: 555 Thirteenth Street, N.W., Suite 701-E
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/142,078
 ; FILING DATE: 10-FEB-1999
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO US97/12652
 ; FILING DATE: 21-JUL-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/762,377
 ; FILING DATE: 06-DEC-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/684,750
 ; FILING DATE: 22-JUL-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Innen, Jeffrey L.
 ; REGISTRATION NUMBER: 28,957
 ; REFERENCE/DOCKET NUMBER: 2314-135, A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-783-6040
 ; TELEFAX: 202-783-6031
 ; INFORMATION FOR SEQ ID NO: 45:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 718 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (cDNA)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Canus geographus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 110..409
 ; US-09-142-078-45

Query Match 56.0%; Score 14; DB 4; Length 718;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 caggtgcacgcacca 15
 ||||||||||||
 DB 118 CAGGTGCATGCCCA 105

RESULT 6
 US-08-420-235B-16/c
 ; Sequence 16, Application US/08420235B
 ; Patent No. 5801042
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 ; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/420,235B
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 45185-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2193 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: N
 ; ANTI-SENSE: N
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..2193
 ; OTHER INFORMATION:
 ; US-08-420-235B-16

Query Match 56.0%; Score 14; DB 1; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgcacat 19
 ||||||||||||
 DB 46 TGCATGCCACGACAT 33

RESULT 7
 US-08-793-624-16/c
 ; Sequence 16, Application US/08793624C
 ; Patent No. 6150093
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
 ; TITLE OF INVENTION: Uses Thereof

FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentl Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 56.0%; Score 14; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgccagcat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 8
PCT-US95-10194-16/c
Sequence 16, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentl Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match 56.0%; Score 14; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgccagcat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 9
PCT-US96-05320A-113/c
Sequence 113, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide sequence of the Haemophilus Influenzae Rd Genome
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 122, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2500
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-113

Query Match 56.0%; Score 14; DB 5; Length 2232;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atccacagcatc 22
|||||
Db 636 ATCCACAGCATTC 623

RESULT 10
US-08-343-101A-5/c
Sequence 5, Application US/08343101A

Patent No. 5830759
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
NUMBER OF INVENTION: Unique Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,101A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-343-101A-5

Query Match 56.0%; Score 14; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattgccagcat 19
|||||
DB 2182 TGCATGCCAGCAT 2169

RESULT 11
US-09-183-688-5/c
Sequence 5, Application US/09183688
Patent No. 6093550
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
NUMBER OF INVENTION: Unique Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-09-183-688-5

Query Match 56.0%; Score 14; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattgccagcat 19
|||||
DB 2182 TGCATGCCAGCAT 2169

RESULT 12
US-08-169-927-1/c
Sequence 1, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
APPLICANT: Dasch, Gregory A
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF INVENTION: Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759

```

; TELEFAX: (301) 295-1022
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Rickettsia prowazekii
; STRAIN: Breinl
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 340..345
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 363..368
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 391..5226
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 379..386
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5270..5306
; PUBLICATION INFORMATION:
; AUTHORS: Carl, M.
; AUTHORS: Dobson, M. E.
; AUTHORS: Ching, W. M.
; AUTHORS: Dasch, G. A.
; TITLE: Characterization of the gene encoding the
; TITLE: protective S-layer protein of Rickettsia
; TITLE: prowazekii, presence of a truncated identical
; TITLE: homolog in Rickettsia typhi
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; DATE: 1990
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

```

```

Query Match          56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 catgcgcagcat 21
|||||
DB 2221 CATGCCAGCATTT 2208

```

```

RESULT 13
US-08-420-235B-1/c
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420.235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-420-235B-1

```

```

Query Match          56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 tgcacgcgcagcat 19
|||||
DB 2138 TGCATCGGCAGCAT 2125

```

```

RESULT 14
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

```

```

Query Match          56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 tgcacgcgcagcat 19
|||||
DB 2138 TGCATCGGCAGCAT 2125

```

```

RESULT 15
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45

```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/10194
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20710 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHEetical: N
;; ANTI-SENSE: N
;; PCT-US95-10194-1

Query Match 56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgacat 19
|||||

Db 2138 TGCATGCCGACAT 2125

RESULT 16
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 52342
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-770-379-18

Query Match 56.0%; Score 14; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgacat 19
|||||

Db 2058 TGCATGCCGACAT 2045

RESULT 17
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18

Query Match 56.0%; Score 14; DB 4; Length 35100;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gtcatgcgcagcatt 19
|||||
Db 2058 TGCAATGCCAGCATT 2045

RESULT 18

US-09-060-756-726
; Sequence 726, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 726
; LENGTH: 304
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-726

Query Match 52.0%; Score 13; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtcatgcgcagcatt 17
|||||
Db 136 gtcatgcgcagcatt 148

RESULT 19

US-09-060-756-570
; Sequence 570, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 570
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match 52.0%; Score 13; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtcatgcgcagcatt 17
|||||
Db 135 gtcatgcgcagcatt 147

RESULT 20

US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgcgcagcatt 20
|||||
Db 406 CATGCCAGCATT 418

RESULT 21

US-08-933-750C-60
; Sequence 60, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga

APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TIVMOR01
CLONE: 140516
US-08-933-750C-60

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ccagcattcgat 25
DB 722 CCAGCATTTCGAT 734

RESULT 22
US-09-224-613-60
Sequence 60, Application US/09234613
Patent No. 6132973
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TIVMOR01
CLONE: 140516
US-09-234-613-60

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ccagcattcgat 25
DB 722 CCAGCATTTCGAT 734

RESULT 23
US-07-862-588B-6
Sequence 6, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch Ieln, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambdiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1624 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match 52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcacgcc 14
|||||
Db 934 CAGGTGCATGCC 946

RESULT 24
US-07-862-588B-5
Sequence 5, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch Jein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambdiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728

REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1775 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..(1625,1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
OTHER INFORMATION: (OTR)
US-07-862-588B-5

Query Match 52.0%; Score 13; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcacgcc 14
|||||
Db 934 CAGGTGCATGCC 946

RESULT 25
US-08-584-226-1/c
Sequence 1, Application US/08584226
Patent No. 5798240
GENERAL INFORMATION:
APPLICANT: Martinus, Susan A.
APPLICANT: Sassanitar, Mandana
APPLICANT: Kim, Sunghoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-TRNA
TITLE OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584,226
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,766
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CP194-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1563
US-08-584-226-1

Query Match 52.0%; Score 13; DB 1; Length 2290;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgcagcagc 14
Db 1233 CAGGTGCTCCGCC 1221

RESULT 26
US-09-134-566-15
Sequence 15, Application US/09134566
Patent No. 5998147
GENERAL INFORMATION:
APPLICANT: Petit, Christine
APPLICANT: Denoyelle-Gryson, Francoise
APPLICANT: Weil, Dominique
APPLICANT: Marlin-Duvernois, Sandrine
TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
FILE REFERENCE: 03495-0175
CURRENT APPLICATION NUMBER: US/09/134,566
CURRENT FILING DATE: 1998-08-14
EARLIER APPLICATION NUMBER: 60/055,863
EARLIER FILING DATE: 1998-08-15
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 15
LENGTH: 2312
TYPE: DNA
ORGANISM: human connexin
US-09-134-566-15

Query Match 52.0%; Score 13; DB 2; Length 2312;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcgcagcagc 22
Db 960 tcgcagcagc 972

RESULT 27
US-09-134-566-8
Sequence 8, Application US/09134566
Patent No. 5998147
GENERAL INFORMATION:
APPLICANT: Petit, Christine
APPLICANT: Denoyelle-Gryson, Francoise
APPLICANT: Weil, Dominique
APPLICANT: Marlin-Duvernois, Sandrine
TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS

TITLE OF INVENTION: HEREDITARY DEFECT
FILE REFERENCE: 03495-0175
CURRENT APPLICATION NUMBER: US/09/134,566
CURRENT FILING DATE: 1998-08-14
EARLIER APPLICATION NUMBER: 60/055,863
EARLIER FILING DATE: 1998-08-15
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 8
LENGTH: 2314
TYPE: DNA
ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcagcagc 22
Db 962 tcgcagcagc 974

RESULT 28
US-08-759-444-2/c
Sequence 2, Application US/08759444
Patent No. 5824309
GENERAL INFORMATION:
APPLICANT: Dassarma, Shilditya
APPLICANT: Morshead, Fazeela
APPLICANT: Stuart, Elizabeth
APPLICANT: Black, Samuel
TITLE OF INVENTION: RECOMBINANT GAS VESICLES AND USES THEREOF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,444
FILING DATE: 05-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,200
FILING DATE: 05-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07880/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8878 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-759-444-2

Query Match 52.0%; Score 13; DB 1; Length 8878;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcagc 17
 |||
 Db 7684 GTGCATCGCCAGC 7672

RESULT 29
 US-08-680-897-1/c
 ; Sequence 1, Application US/08680897
 ; Patent No. 6008051
 ; GENERAL INFORMATION:
 ; APPLICANT: Dassarma, Shiladitya
 ; APPLICANT: Ng, Mai-lap
 ; TITLE OF INVENTION: RECOMBINANT VECTOR AND PROCESS FOR CELL
 ; TITLE OF INVENTION: FLOTATION
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/680,897
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/271,270
 ; FILING DATE:
 ; APPLICATION NUMBER: 944,581
 ; FILING DATE: 14-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digilio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 8680
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9880 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-680-897-1

Query Match 52.0%; Score 13; DB 3; Length 9880;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcagc 17
 |||
 Db 8684 GTGCATCGCCAGC 8672

RESULT 30
 US-08-851-843A-92/c
 ; Sequence 92, Application US/08851843A
 ; Patent No. 6093809
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Harley, Calvin H.
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: NO. 6093809el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0300
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 92:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgcgcagcat 19
 |||
 Db 19 CATGCAGCAGCAT 8

RESULT 31
 US-08-974-549A-384/c
 ; Sequence 384, Application US/08974549A
 ; Patent No. 6166178
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.
 ; APPLICANT: Harley, Calvin H.
 ; APPLICANT: Andrews, William H.
 TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/911,312
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/912,951
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/915,503
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17618
 FILING DATE: 01-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17885
 FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002610US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 384:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 FEATURE:
 NAME/KEY:
 LOCATION: 1..19
 OTHER INFORMATION: /note="TCPI.6 primer"
 US-08-974-549A-384

Query Match 48.0%; Score 12; DB 4; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catgccagcat, 19

Db 19 CATGCCAGCAT 8

RESULT 32
 US-08-373-124A-1692/C
 ; Sequence 1692, Application US/08373124A
 ; Patent No. 5646042
 ; GENERAL INFORMATION:
 ; APPLICANT: Stinchcomb, Dan T.
 ; APPLICANT: Draper, Kenneth
 ; APPLICANT: McSwigen, James
 ; APPLICANT: Jarvis, Thale
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
 ; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
 ; TITLE OF INVENTION: CANCER USING RIBOZYMES
 ; NUMBER OF SEQUENCES: 2627
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/373,124A
 ; FILING DATE: January 13, 1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/245,466
 ; FILING DATE: May 18, 1994
 ; APPLICATION NUMBER: 08/192,943
 ; FILING DATE: February 7, 1994
 ; APPLICATION NUMBER: 07/987,132
 ; FILING DATE: December 7, 1992
 ; APPLICATION NUMBER: 07/936,422
 ; FILING DATE: August 26, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wardburg, Richard
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 209/035
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 1692:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 38 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-373-124A-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 gccagcatcg 23
 Db 38 GCCAGCATTCG 27

RESULT 33
 US-08-435-628-1692/C
 ; Sequence 1692, Application US/08435628
 ; Patent No. 5817796

```

; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwigen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: California
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,628
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,124
; FILING DATE: January 13, 1995
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Waiburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELETYPE: 67-3510
; INFORMATION FOR SEQ ID NO: 1692:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-435-628-1692

Query Match      48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 gccagatttcg 23
      |||
DB      38 gccagatttcg 27

RESULT 34
US-09-023-228B-59
; Sequence 59, Application US/09023228B
; Patent No. 6140490
; GENERAL INFORMATION:
; APPLICANT: BIESSECKER, GREGORY
; APPLICANT: GOLD, LARRY
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
; TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS

```

```

; NUMBER OF SEQUENCES: 157
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Place #200
; CITY: Denver
; STATE: Colorado
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,228B
; FILING DATE: 12-FEBRUARY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01739
; FILING DATE: 30 JAN 1997
; APPLICATION NUMBER: 08/595,335
; FILING DATE: 1 FEB 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson, Esq.
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX50/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
; US-09-023-228B-59

Query Match      48.0%; Score 12; DB 3; Length 61;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      14 gccatttcgat 25
      |||
DB      23 CAGCAUUCGAU 34

RESULT 35
US-08-488-402A-16/C
; Sequence 16, Application US/08488402A
; Patent No. 5837456
; GENERAL INFORMATION:
; APPLICANT: GOLD ET AL.
; TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO
; TITLE OF INVENTION: CHORIONIC GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN
; TITLE OF INVENTION: HORMONES
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/488,402A
FILING DATE: 07 JUNE 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-488-402A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 aggtgcacgcc 14
|||||
Db 25 AGTGCGATCGCC 14

RESULT 36
US-08-484-552A-16/c
Sequence 16, Application US/08484552A
Patent No. 5849890
GENERAL INFORMATION:
APPLICANT: GOLD, LARRY
APPLICANT: JAYASENA, SUMEDHA
APPLICANT: NIEMULANDT, DAN
APPLICANT: DAVIS, KEN
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO CHORIONIC
TITLE OF INVENTION: GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,552A

FILING DATE: 07 JUNE 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-484-552A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 aggtgcacgcc 14
|||||
Db 25 AGTGCGATCGCC 14

RESULT 37
PCT-US96-09472-16/c
Sequence 16, Application PC/TUS9609472
GENERAL INFORMATION:
APPLICANT: GOLD ET AL.
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE
TITLE OF INVENTION: LIGANDS TO CHORIONIC GONADOTROPIN HORMONE AND RELATED
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09472
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/488,402
APPLICATION NUMBER:

FILING DATE: 07 JUNE 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,552
FILING DATE: 07 JUNE 1995
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
PCT-US96-09472-16

Query Match 48.0%; Score 12; DB 5; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 aggtcgcgcgc 14
|||||
Db 25 AGTGCATCGCC 14

RESULT 38
US-08-998-416-93/c
Sequence 93, Application US/0898416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1031UP
US-08-998-416-93

Query Match 48.0%; Score 12; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tgcacgcgcgc 17
|||||
Db 292 TGCATCGCCGAC 281

RESULT 39
US-08-743-637B-184/c
Sequence 184, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUBLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 438 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
US-08-743-637B-184

Query Match 48.0%; Score 12; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcg 12
|||||
DB 28 GCAGGTGCATCG 17

RESULT 40

US-08-998-416-56/c

Sequence 56, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mont, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtle, Philipp

APPLICANT: Reibischung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 656 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: PAG1016RP

US-08-998-416-56

Query Match 48.0%; Score 12; DB 4; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattgcacgc 17
|||||
DB 53 TGCATTGCCACGC 42

RESULT 41

US-08-743-637B-175/c

Sequence 175, Application US/08743637B

Patent No. 5994066

GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.

APPLICANT: PICARD, Francois J.

APPLICANT: OUELLETTE, Marc

APPLICANT: ROY, Paul H.

TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA

TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND

TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED

TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...

NUMBER OF SEQUENCES: 273

CORRESPONDENCE ADDRESS:

ADDRESSEE: QUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE

CITY: MILWAUKEE

STATE: WISCONSIN

COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743,637B

FILING DATE: 04-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,840

FILING DATE: 11-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BAKER, Jean C.

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586, 90012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:

LENGTH: 660 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-743-637B-175

Query Match 48.0%; Score 12; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cagatttcgat 25
|||||
DB 627 CAGATTTCGAT 616

RESULT 42

US-08-526-840B-175/c

Sequence 175, Application US/08526840B

Patent No. 6001564

GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.

APPLICANT: OUELLETTE, Marc

APPLICANT: ROY, Paul H.

TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND

TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY

TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES

TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:

ADDRESSEE: QUARLES & BRADY

STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 660 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-526-840B-175

Query Match 48.0%; Score 12; DB 3; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cagcattcgat 25
|||||
DB 627 CAGCATTCGAT 616

RESULT 43
US-08-998-416-767
Sequence 767, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBVA GOSSYPIL
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 767:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1495RP
US-08-998-416-767

Query Match 48.0%; Score 12; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgccagcattc 21
|||||
DB 356 TCGCCAGCATTT 367

RESULT 44
US-08-998-416-237/C
Sequence 237, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBVA GOSSYPIL
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

```

; INFORMATION FOR SEQ ID NO: 237:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1215RP
; US-08-998-416-237

```

```

Query Match          48.0%; Score 12; DB 4; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2  caggtcgcgcgc 13
          |||
Db      461  CAGGTGCATCGC 450

```

```

RESULT 45
US-08-990-823-39
; Sequence 39, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 39
; LENGTH: 841
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: Modified base
; OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-39

```

```

Query Match          48.0%; Score 12; DB 4; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      7  gcatagcgcagca 18
          |||
Db      370  gcatagcgcagca 381

```

```

Search completed: October 9, 2001, 15:55:36
Job time: 13212 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:31 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196F-6

Perfect score: 25
Sequence: 1 gcagtgatcgcagcagatcgcagat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 443280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	17	68.0	498	21	AI532390	AI532390 SD03891.5
C 2	16	68.0	376	171	BF933544	BF933544 CMI-NT027
C 3	16	64.0	399	156	D24778	D24778 R1CR2545A.R
C 4	16	64.0	429	107	A0030894	A0030894 A0030894
C 5	16	64.0	510	107	A0031475	A0031475 A0031475
C 6	16	64.0	548	258	P968R	AI390674 Leishman1
C 7	16	64.0	666	107	A0030895	A0030895 A0030895
C 8	16	64.0	704	107	A0078750	A0078750 A0078750
C 9	16	64.0	751	170	BF864751	BF864751 963054E05
C 10	15	60.0	258	246	A578088	A578088 18f02_Sho
C 11	15	60.0	266	19	AI410504	AI410504 EST338797
C 12	15	60.0	338	162	BE065518	BE065518 RC3-BT031
C 13	15	60.0	341	31	AV551712	AV551712 EST511712
C 14	15	60.0	380	12	AA800788	AA800788 EST190285
C 15	15	60.0	393	158	HE0965	HE0965 Y722C07.r1
C 16	15	60.0	394	189	T86053	T86053 Y662405.r1
C 17	15	60.0	396	167	BE429060	BE429060 MTD013.H0
C 18	15	60.0	399	162	BE044431	BE044431 hc45d04.x
C 19	15	60.0	414	18	AI268648	AI268648 q03f05.x
C 20	15	60.0	435	223	AG027122	AG027122 C1T-HSP-2
C 21	15	60.0	437	158	HE92800	HE92800 Y795f10.r1
C 22	15	60.0	438	187	N92081	N92081 za23h03.r1
C 23	15	60.0	445	32	AV722315	AV722315 AV722315
C 24	15	60.0	453	108	A0166733	A0166733 A0166733
C 25	15	60.0	463	167	BE446670	BE446670 WHE1139_F
C 26	15	60.0	466	1	AA026650	AA026650 z199e08.r
C 27	15	60.0	475	12	AA821080	AA821080 GM05622.5
C 28	15	60.0	475	158	HE0979	HE0979 Y722f01.r1
C 29	15	60.0	475	173	BE101817	BE101817 RH122_22
C 30	15	60.0	491	23	AI692369	AI692369 wd63f02.x
C 31	15	60.0	491	146	BF211097	BF211097 601812706
C 32	15	60.0	499	171	BF916376	BF916376 IL3-UT011
C 33	15	60.0	505	156	C98143	C98143 C98143 Rice
C 34	15	60.0	507	189	T84279	T84279 Y447D04.r1
C 35	15	60.0	514	154	BE467085	BE467085 IA04C06.B
C 36	15	60.0	524	104	AI957283	AI957283 ul85e12.x
C 37	15	60.0	528	158	HE74120	HE74120 vs15f08.r1
C 38	15	60.0	536	174	BE623022	BE623022 na635e06
C 39	15	60.0	539	170	BF839030	BF839030 PM2-HT035
C 40	15	60.0	545	237	A2071920	A2071920 RPCI-23-3
C 41	15	60.0	547	118	AMS88900	AMS88900 ra06g05.Y
C 42	15	60.0	553	241	AZ283519	AZ283519 RPCI-23-1
C 43	15	60.0	567	156	C98142	C98142 C98142 Rice
C 44	15	60.0	581	7	AA405993	AA405993 zu66f09.r
C 45	15	60.0	605	233	AO783596	AO783596 HS_3121_A

ALIGNMENTS

RESULT 1
AI532390/c 498 bp mRNA EST 18-MAR-1999
LOCUS SD03891.5prlme SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION POT2 Drosophila melanogaster cDNA clone SD03891 5prlme, mRNA
sequence.

ACCESSION AI532390
VERSION AI532390.1 GI:4446525
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
AUTHORS P., Lewis, S. and Rubin, G.M.

TITLE BDGP/HIMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3300, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, estfruitfly.berkeley.edu
Plate: 38 row: H column: 7
High quality sequence stop: 433.

FEATURES
source
1..498
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD03891"
/cloned="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pot2; Site1: EcoRI; Site2: XhoI; sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."

BASE COUNT 158 a 106 c 124 g 110 t
ORIGIN

Query Match 68.0%; Score 17; DB 21; Length 498;
Best local Similarity 100.0%; Pred. No. 4.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcacgcacgacatccg 23
|||||
Db 266 GCATCCGCACGATTCG 250

RESULT 2
BF933544/c 376 bp mRNA EST 22-JAN-2001
LOCUS CMI-NT0270-211200-661-g03 NT0270 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF933544
VERSION BF933544.1 GI:12350868
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?rl=CMI2-CMI-NT0270-211200-661-g03&rl=2000-12-21&rl=1>)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.
Location/Qualifiers
1..376

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0270"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT      77 a      104 c      99 g      96 t
ORIGIN

Query Match      64.0%; Score 16; DB 171; Length 376;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggtgcacgcacgcat 19
|||||
Db 329 GGTGATCGCCAGCAT 314

RESULT 3
D24778 399 bp mRNA EST 08-JUL-1999
DEFINITION RICE2545A Rice root Oryza sativa cDNA clone R2545_1A, mRNA
sequence.
ACCESSION D24778
VERSION D24778.1 GI:428626
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 399)
REFERENCE
AUTHORS Minobe,Y. and Sasaki,T.
TITLE Rice cDNA from root
JOURNAL Unpublished (1995)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT="RGP".

FEATURES
source location/Qualifiers
1..399
/organism="Oryza sativa"
/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="R2545_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 114 a 70 c 102 g 112 t 1 others
ORIGIN

Query Match      64.0%; Score 16; DB 156; Length 399;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggtgcacgcacgcat 19
|||||
Db 251 GGTGATCGCCAGCAT 266

RESULT 4

```

```

AU030894/c
LOCUS AU030894 429 bp mRNA EST 29-OCT-1998
DEFINITION AU030894 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E60405_1A, mRNA sequence.
ACCESSION AU030894
VERSION AU030894.1 GI:376784
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 429)
REFERENCE
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from immature leaf including apical meristem
JOURNAL Unpublished (1997)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT="RGP".

FEATURES
source location/Qualifiers
1..429
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E60405_1A"
/clone_lib="Rice cDNA from immature leaf including apical
meristem"
/dev_stage="Immature"
/note="Organ: leaf; Immature leaf including apical
meristem (under long day condition)."
BASE COUNT 106 a 122 c 101 g 99 t 1 others
ORIGIN

Query Match      64.0%; Score 16; DB 107; Length 429;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 catgcacgcatctg 23
|||||
Db 302 CATGCCAGCATTCG 287

RESULT 5
AU031475/c
LOCUS AU031475 510 bp mRNA EST 30-OCT-1998
DEFINITION AU031475 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E61659_2Z, mRNA sequence.
ACCESSION AU031475
VERSION AU031475.1 GI:3767365
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 510)
REFERENCE
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from immature leaf including apical meristem
JOURNAL Unpublished (1997)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305

```

Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasakileabr@affrc.go.jp

PROJECT = 'RGP'

FEATURES

SOURCE

Location/Qualifiers
1. 510
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E61659_22"
/clone_11b="Rice cDNA from immature leaf including apical meristem"
/dev_stage="Immature"
/note="Organ: Leaf; Immature leaf including apical meristem (under long day condition)"

BASE COUNT 176 a 98 c 99 g 137 t
ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 510;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcattcg 23
|||||
Db 136 CATGCCAGCATTTCG 121

RESULT 6 548 bp DNA GSS 25-JUL-2000
LOCUS p968R
DEFINITION Leishmania major Friedlin PAC p968 right end-sequence, genomic survey sequence.

ACCESSION AL390674
VERSION AL390674.1 GI:9501650
KEYWORDS GSS.

SOURCE Leishmania major.
ORGANISM Leishmania major.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 548)
Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F.

TITLE A physical map of the Leishmania major Friedlin genome
JOURNAL Genome Res. 8 (2), 135-145 (1998)
MEDLINE 98146435
REFERENCE 2 (bases 1 to 548)

AUTHORS Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.

COMMENT Direct Submission
Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allicates@sanger.ac.uk

see <http://www.ebi.ac.uk/parasites/leish.html>

DETAILS of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/L_major/
The primer sequence can be obtained from allicates@sanger.ac.uk.

FEATURES Location/Qualifiers
1. 548
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="PAC p968"

BASE COUNT 86 a 167 c 163 g 132 t
ORIGIN

Query Match 64.0%; Score 16; DB 258; Length 548;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcagcatt 20

Db 541 GTGCATGCCAGCATT 526
|||||

RESULT 7 666 bp mRNA EST 29-OCT-1998
LOCUS AU030895/c
DEFINITION AU030895 Rice cDNA from immature leaf including apical meristem

Oryza sativa cDNA clone E60405_22, mRNA sequence.
AU030895
AU030895.1 GI:3766785
EST.

SOURCE Oryza sativa.
ORIGIN Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 666)
Sasaki, T. and Yamamoto, K.

Rice cDNA from immature leaf including apical meristem
Unpublished (1997)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasakileabr@affrc.go.jp

PROJECT = 'RGP'

FEATURES

SOURCE

Location/Qualifiers
1. 666
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E60405_22"
/clone_11b="Rice cDNA from immature leaf including apical meristem"
/dev_stage="Immature"
/note="Organ: Leaf; Immature leaf including apical meristem (under long day condition)"

BASE COUNT 195 a 149 c 131 g 181 t 10 others
ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 666;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcattcg 23
|||||
Db 293 CATGCCAGCATTTCG 278

RESULT 8 704 bp mRNA EST 18-OCT-1999
LOCUS AU078750/c
DEFINITION AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone E0723_42, mRNA sequence.

ACCESSION AU078750
VERSION AU078750.1 GI:6062509
KEYWORDS EST.
SOURCE Oryza sativa.

ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 704)
Sasaki, T. and Yamamoto, K.
Rice cDNA from panicle at flowering stage
Unpublished (1996)
Contact: Takuji Sasaki

National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305

Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasak@abr.affrc.go.jp

PROJECT = "RGP".

FEATURES

Source

Location/Qualifiers
1..704
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E0723.42"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"
BASE COUNT 203 a 167 c 142 g 191 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 704;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcttcg 23
|||||

Db 346 CATGCCAGCATTCG 331

RESULT 9
BF864751/c 761 bp mRNA EST 19-JAN-2001
LOCUS

DEFINITION 963054E05.y1 C. reinhardtii CC-1690, stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF864751 GI:12254895

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

DUKE UNIVERSITY

TELEPHONE

FAX

EMAIL

LOCATION/QUALIFIERS

1..761

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into Lambda Zap II

(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 179 c 308 g 117 t 4 others
ORIGIN

Query Match 64.0%; Score 16; DB 170; Length 761;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 aggtgcacgcgcagca 18
|||||

Db 466 AGGTGCATCGCCAGCA 451

RESULT 10
AZ578088/c 258 bp DNA GSS 08-DEC-2000
LOCUS

DEFINITION 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
sp. NGR234 genomic clone 18f02, DNA sequence.

ACCESSION AZ578088 GI:11605015

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

UNIVERSITY OF GENEVA

1 Chemin de l'Imperatrice, Chambesey/Geneva 1292, Switzerland

Tel: +44(0)1603450000

Fax: +44(0)1603450045

Email: virginie.viprey@bsrc.ac.uk

Class: shotgun

LOCATION/QUALIFIERS

1..258

/organism="Rhizobium sp. NGR234"

/strain="ANU265"

/db_xref="taxon:394"

/clone="18f02"

/clone_lib="Shot-gun genomic library of Rhizobium strain
ANU265"

/note="Vector: M13; derivative strain of NGR234 cured of
pNGR234a"

BASE COUNT 51 a 82 c 79 g 46 t

ORIGIN

Query Match 60.0%; Score 15; DB 246; Length 258;
Best Local Similarity 100.0%; Pred. No. 72;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcgcagcattcga 24
|||||

Db 175 TCGCCAGCATTCGA 161

RESULT 11
A1410504

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

EST.

SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 266)
AUTHORS Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source
1..266
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone_lib="RHEC29"
/note="Organ: heart; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 56 a 61 c 64 g 85 t

ORIGIN

Query Match 60.0%; Score 15; DB 19; Length 266;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggtgcatgcgcagca 18
|||||

Db 90 ggtgcatgcgcagca 104

RESULT 12
BE065518/c 338 bp mRNA EST 09-JUN-2000
LOCUS
DEFINITION RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE065518
VERSION BE065518.1 GI:8410168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 338)
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-BT0316-170200-014-b06&t3=2000-02-17&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 338.
Location/Qualifiers
1..338
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0316"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 76 a 100 c 66 g 96 t

ORIGIN

Query Match 60.0%; Score 15; DB 162; Length 338;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcagcatttc 22
|||||

Db 97 catgcgcagcatttc 83

RESULT 13
AV551712 341 bp mRNA EST 06-SEP-2000
LOCUS
DEFINITION AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZ13E05R 5', mRNA sequence.
ACCESSION AV551712
VERSION AV551712.1 GI:8723125
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 341)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..341
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="RZ13E05R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 73 a 85 c 53 g 130 t

ORIGIN

Query Match 60.0%; Score 15; DB 31; Length 341;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcgcagcatttcga 24

Insert Length: 1146 Std Error: 0.00
Seq primer: M13Rp1
High quality sequence stop: 180.

FEATURES

source

1. .394

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:112856"

/clone_lib="Soares fetal liver spleen 1NPLS"

/sex="male"

/dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGAGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3']"

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 86 a 98 c 82 g 122 t 6 others
ORIGIN

Query Match 60.0%; Score 15; DB 189; Length 394;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
|||||

Db 183 CATGCCAGCATTTTC 169

RESULT 17
BE429060 396 bp mRNA EST 26-JUL-2000
LOCUS BE429060/c

DEFINITION MTD013.H06p990621 ITCC MTD Durum Wheat Root Library Triticum
turgidum subsp. durum cDNA clone MTD013.H06, mRNA sequence.

ACCESSION BE429060
VERSION BE429060.1 GI:9426903
KEYWORDS EST.

SOURCE durum wheat.
ORGANISM Triticum turgidum subsp. durum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 396)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Hermann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pechonin, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M. and Wenzel, G.

International Triticace EST Cooperative (ITCC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)

JOURNAL COMMENT Contact: Joudrier P
INRA, unite de Biochimie et Biologie Molculaire des Cereales
2, place VIALA, 34060 Montpellier cedex 01 FRANCE
Tel: 33 4 99 61 23 48
Fax: 33 4 99 61 23 84

FEATURES Email: joudrierensam.inra.fr
International Triticace EST Cooperative (ITCC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers

1. .396
/organism="Triticum turgidum subsp. durum"
/cultivar="Siliiana"
/db_xref="taxon:4567"
/clone="MTD013.H06"
/clone_lib="ITCC MTD Durum Wheat Root Library"

/tissue_type="root"
/dev_stage="3-day-old seedling, water-stressed"
/note="Vector: pSPORT1; T7 primers used. See pSPORT1
polylinker site. 0.3-2.0 kbp average insert size."

BASE COUNT 96 a 87 c 116 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 396;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcagtgatcgcgca 15
|||||

Db 75 GCAGTGATCGCGCA 61

RESULT 18
BE044431 399 bp mRNA EST 08-JUN-2000
LOCUS BE044431

DEFINITION h045d04.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
IMAGE:3040327 3', mRNA sequence.

ACCESSION BE044431
VERSION BE044431.1 GI:8361484
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 399)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco.
Location/Qualifiers

FEATURES

source

1. .399

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3040327"

/clone_lib="Soares_NFL.T.GBC.S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NT7, and B-cell
NCI CGAP, GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 91 a 94 c 96 g 118 t
ORIGIN

Query Match 60.0%; Score 15; DB 162; Length 399;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
|||||

Db 329 CATGCCAGCATTTTC 343

RESULT 19
A1268648

LOCUS A1268648 414 bp mRNA EST 17-NOV-1998
 DEFINITION GO39f05.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1910913 3',
 mRNA sequence.
 ACCESSION A1268648
 VERSION A1268648.1 GI:3887815
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.linnl.gov/bbrp/image/image.html
 Seq primer: -400p from GIBCO
 High quality sequence stop: 398.
 Location/Qualifiers
 1..414
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1910913"
 /clone_1id="NCI_CGAP_Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 neuroendocrine lung carcinoma, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 122 a 82 c 109 g 101 t
 ORIGIN

Query Match 60.0%; Score 15; DB 18; Length 414;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 ||||||||||||
 Db 353 catgccagcatttc 367

RESULT 20
 LOCUS A0027122 435 bp DNA GSS 30-JUN-1998
 DEFINITION CIT-HSP-2323H13.TF CIT-HSP Homo sapiens genomic clone 2323H13, DNA
 sequence.
 ACCESSION A0027122
 VERSION A0027122.1 GI:3267344
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 435)
 AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
 Simon, M., and Venter, J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map

JOURNAL Building (1998)
 COMMENT Unpublished (1998)
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2323H13"
 /clone_1id="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
 HindIII"

BASE COUNT 140 a 97 c 77 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 223; Length 435;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 ||||||||||||
 Db 178 catgccagcatttc 192

RESULT 21
 LOCUS H69200/c 437 bp mRNA EST 19-OCT-1995
 DEFINITION YR59f10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:213067 5', mRNA sequence.
 ACCESSION H69200
 VERSION H69200.1 GI:1030526
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 437)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissole, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
 , M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
 , B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

TITLE
 JOURNAL
 MEDLINE
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1098
 High quality sequence stops: 300
 Source: IMAGE Consortium, LINL
 This clone is available royalty-free through LINL; contact the
 IMAGE Consortium (info@image.linnl.gov) for further information.
 Insert length: 1098 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 300.

FEATURES
source

Location/Qualifiers
1. .437

/organism="Homo sapiens"
/db_xref="GDB:3778371"

/db_xref="taxon:9606"
/clone_image="IMAGE:213067"

/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"

/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pTT3D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I...oligo(dT) primer [5' AACGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 97 a 113 c 89 g 134 t 4 others
ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 437;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
DB 139 CATGCCAGCATTC 125

RESULT 22 438 bp mRNA EST 04-APR-1996
LOCUS N92081
DEFINITION za23h03.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:293429 5', mRNA sequence.

ACCESSION N92081
VERSION N92081.1 GI:1264390

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 438)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston

, R., Williamson, A., Woldmann, P. and Wilson, R.
The WashU-Merck EST Project

TITLE Unpublished (1995)
JOURNAL Contact: Wilson RK

COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: estewatson.wustl.edu
This clone is available royalty-free through INFL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET

High quality sequence stop: 321.
Location/Qualifiers

1. .438
/organism="Homo sapiens"

/db_xref="GDB:3801206"
/db_xref="taxon:9606"

/clone_image="IMAGE:293429"
/clone_lib="Soares fetal liver spleen INFLS"

/sex="male"
/dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTT3D (Pharmacia)

with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTT3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 104 a 111 c 101 g 118 t 4 others
ORIGIN

Query Match 60.0%; Score 15; DB 187; Length 438;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
DB 308 CATGCCAGCATTC 294

RESULT 23 445 bp mRNA EST 16-OCT-2000
LOCUS AV722315/c

DEFINITION AV722315 HTB Homo sapiens cDNA clone HTBAMG01 5', mRNA sequence.

ACCESSION AV722315
VERSION AV722315.1 GI:10824678

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 445)

REFERENCE Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu

, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.

Homo sapiens cDNA HTB clones
Unpublished (2000)

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhanjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China
Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers
1. .445

/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone_image="HTBAMG01"
/clone_lib="HTB"

/tissue_type="Hypothalamus"
/dev_stage="Adult"

/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 102 a 123 c 104 g 116 t
ORIGIN

Query Match 60.0%; Score 15; DB 32; Length 445;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
DB 296 CATGCCAGCATTC 282

RESULT 24 453 bp mRNA EST 23-JAN-2001
LOCUS AU166733/c

DEFINITION AU166733 Rice callus *Oryza sativa* subsp. *japonica* cDNA clone C50066
, mRNA sequence.
ACCESSION AU166733
VERSION AU166733.1 GI:12405132
KEYWORDS EST.
SOURCE *Oryza sativa* subsp. *japonica*.
ORGANISM *Oryza sativa* subsp. *japonica*.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ethnartoideae; Oryzaceae; *Oryza*.
REFERENCE 1 (bases 1 to 453)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from callus (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
C50066_98Z.

FEATURES
source
1. 453
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C50066"
/note="lib="Rice callus"
/note="Vector: Bluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of Bluescript II SK+ phagemid."
BASE COUNT 164 a 83 c 140 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 108; Length 453;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgacat 20
|||||
Db 349 tgcacgcacgacat 335

RESULT 25
BE446670/c 463 bp mRNA EST 25-JUL-2000
LOCUS WHE1139_F07_L1325 Wheat etiolated seedling root normalized cDNA
DEFINITION library Triticum aestivum cDNA clone WHE1139_F07_L13, mRNA
sequence.
ACCESSION BE446670
VERSION BE446670.1 GI:9446232
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
Triticeae; Triticum.
REFERENCE 1 (bases 1 to 463)
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han,
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,
Rausch, C.J., Seaton, C.L., Tong, J.C., and Zhang, D.
TITLE The structure and function of the expressed portion of the wheat
JOURNAL genomes - Normalized root cDNA library
COMMENT Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragene SK primer.
Location/Qualifiers
1. 463
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1139_F07_L13"
/clone.lib="Wheat etiolated seedling root normalized cDNA
library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli DH10b"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid
Bluescript SK, Site_1: EcoRI; Site_2: XhoI; Seeds were
surface-sterilized, germinated and grown aseptically in
the dark at room temperature on filter paper with water,
nystatin and cefotaxime in covered crystallization
dishes. Roots were harvested. The tissue, total RNA, and
poly(A) RNA were prepared, a cDNA library was made in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. The cDNA clones were in vivo
excised to give phagescript phagemids before
normalization was carried out. The mass excision of
phagemid library and normalization were done in HT Nguyen
lab by D. Zhang at Texas Tech University. Normalization
protocol used was that of Soares. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT 119 a 98 c 132 g 114 t
ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 463;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaagtgacgacca 15
|||||
Db 111 gcaagtgacgacca 97

RESULT 26
AA026650/c 466 bp mRNA EST 09-MAY-1997
LOCUS 2399e08_r1 Soares.pregnant-uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:469190 5', mRNA sequence.
ACCESSION AA026650
VERSION AA026650.1 GI:1492715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacey, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
Underwood, K., Woldmann, P., Waterston, R., Wilson, R., and Marra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1129 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 364.
 Location/Qualifiers

FEATURES

Source
 1. 466
 /organism="Homo sapiens"
 /db_xref="GDB:3755192"
 /db_xref="taxon:9606"
 /clone="IMAGE:469190"
 /clone_lib="Soares_pregnant_uterus_NbHPV"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site: 1; Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5',
 AACTGAGAGATTCGGCGCCGCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."
 BASE COUNT 114 a 124 c 100 g 128 t
 ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 466;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 catgccagcattc 22
 |||||||
 Db 263 catgccagcattc 249

RESULT 27
 AA821080 475 bp mRNA EST 25-NOV-1998
 LOCUS
 DEFINITION GM096622 5prime GM Drosophila melanogaster ovary Bluescript
 Drosophila melanogaster cDNA clone GM096622 5prime, mRNA sequence.
 ACCESSION AA821080
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidae; Drosophilidae; Drosophila.
 1 (bases 1 to 475)
 Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
 P., Lewis, S. and Rubin, G.M.
 BDCP/HHMI Drosophila EST Project
 Unpublished (1997)
 Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 USA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Plate: 96 row: B column: 10
 High quality sequence stop: 439.
 Location/Qualifiers

FEATURES
 Source
 1. 475
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="GM096622"
 /clone_lib="GM Drosophila melanogaster ovary Bluescript"
 /sex="female"
 /dev_stage="newly eclosed females: germlinum-stage 6"

/lab_host="SOLR"
 /note="Organ: ovary; Vector: Bluescript SK; Site: 1; EcoRI;
 Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
 synthesis kit. Oligo dt-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"
 BASE COUNT 128 a 147 c 121 g 79 t
 ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 aggtgcagccagc 17
 |||||||
 Db 344 AGGTGCATCCGACG 358

RESULT 28
 H60979/c 475 bp mRNA EST 06-OCT-1995
 LOCUS
 DEFINITION Y122101.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:206041 5', mRNA sequence.
 ACCESSION H60979
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
 R., Williamson, A., Woldmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1101
 High quality sequence stops: 339
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1101 Std Error: 0.00
 Seq primer: M13RPI
 High quality sequence stop: 339.
 Location/Qualifiers

FEATURES

Source
 1. 475
 /organism="Homo sapiens"
 /db_xref="GDB:3775172"
 /db_xref="taxon:9606"
 /clone="IMAGE:206041"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site: 1; Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5', AACTGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 109 a 123 c 95 g 143 t 5 others
 ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 |||||||

DB 119 CATGCCAGCATTTC 105

RESULT 29
 BG101817 475 bp mRNA EST 30-JAN-2001
 LOCUS RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
 DEFINITION
 ACCESSION BG101817
 VERSION BG101817.1 GI:12616762
 KEYWORDS EST.
 SOURCE Sorghum prolinguam.
 ORGANISM Sorghum prolinguam.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 I (bases 1 to 475)
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
 , L.H.
 An EST database from Sorghum: Sorghum prolinguam rhizomes
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpatt@uga.edu
 Seq primer: POLYTMix
 High quality sequence start: 11
 High quality sequence stop: 408
 POLYA-No.

FEATURES
 Location/Qualifiers
 1..475
 /organism="Sorghum prolinguam"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH122)"
 /note="Organ: Rhizomes; Vector: pBluescript II from lambda
 zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 128 a 107 c 113 g 127 t

Query Match 60.0%; Score 15; DB 173; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggcgcatgcga 15
 |||||||

DB 240 GCAGGTGCATGCCA 254

RESULT 30
 AI692369 491 bp mRNA EST 17-DEC-1999
 LOCUS wd63f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3',
 DEFINITION mRNA sequence.
 ACCESSION AI692369
 VERSION AI692369.1 GI:4969709
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 491)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 569 Std Error: 0.00
 Seq primer: 40UP from Gibco
 High quality sequence stop: 471.

FEATURES
 Location/Qualifiers
 1..491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:2336283"
 /clone_lib="NCI-CGAP_Lu24"
 /rname_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI-CGAP Lu5 was prepared and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (clones
 141920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 151 a 94 c 129 g 116 t 1 others

Query Match 60.0%; Score 15; DB 23; Length 491;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 |||||||

DB 344 CATGCCAGCATTTC 358

RESULT 31
 BF211097 491 bp mRNA EST 06-NOV-2000
 LOCUS BF211097/c
 DEFINITION 601812706f1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4047120 5',
 mRNA sequence.
 ACCESSION BF211097
 VERSION BF211097.1 GI:11104683
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 491)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1CM875 row: 1 column: 01
 High quality sequence start: 56
 High quality sequence stop: 491.

FEATURES

source

1. 491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4047120"
 /clone_lib="NIH-MGC-54"
 /tissue_type="from chronic myelogenous leukemia"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggcgccgcgcgc); Site_2: SfiI (ggccatggcc
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-ATCTAGAGCGCGAGCGCGCATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 123 a 128 c 119 g 121 t

ORIGIN

Query Match 60.0%; Score 15; DB 146; Length 491;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgagcagcatttc 22

|||||

Db 305 CATGCCAGCATTC 291

RESULT 32

BF916376

LOCUS

IL3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BF916376.1 GI:12307834

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 499)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zagro,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0115-

11200-377-F05&t3=2000-12-11&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 496.

Location/Qualifiers

1. 499

FEATURES

source

Query Match 60.0%; Score 15; DB 156; Length 505;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0115"
 /dev_stage="Adult"

/note="Organ: uterus, tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 121 a 121 c 114 g 143 t

ORIGIN

Query Match 60.0%; Score 15; DB 171; Length 499;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaagtgcagccca 15

|||||

Db 216 GCAGGTGCATGCCCA 230

RESULT 33

C98143/c

LOCUS

C98143 Rice callus Oryza sativa subsp. japonica cDNA clone C0777_7A

DEFINITION

, mRNA sequence.

ACCESSION

C98143 D28202

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

Oryza sativa subsp. japonica.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzoae; Oryza.

1 (bases 1 to 505)

Yamamoto,K. and Sasaki,T.

Rice cDNA from callus (1998)

Unpublished (1998)

On Oct 19, 1998 this sequence version replaced gi:454479.

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasakia@affrc.go.jp

D28202: Submitted (01-JAN-1994)

PROJECT="RG"

POLYA=No.

Location/Qualifiers

1. 505

/organism="Oryza sativa subsp. japonica"

/strain="cultivar Nipponbare, sub-species japonica"

/db_xref="taxon:39947"

/clone_lib="C0777_7A"

/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:

oligo(dT) as a primer and ligating to the SalI/NotI site

of pBluescript II SK+ phagemid.

2 others

BASE COUNT 158 a 103 c 170 g 72 t

ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 505;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	6	tgcctccagcatt	20
Db	475	TGCATGCCAGCATTT	461
RESULT	34		
LOCUS	T84279/c		
DEFINITION	Yd47B04.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:111343 5', mRNA sequence.		
ACCESSION	T84279		
VERSION	T84279.1	GI:712567	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 507) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,R.M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished (1995)		
TITLE	Contact: Wilson RK		
JOURNAL	Washington University School of Medicine		
COMMENT	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Insert Size: 1225 High quality sequence stops: 283 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1225 Std Error: 0.00 Seq primer: M13Rp1 High quality sequence stop: 283. Location/Qualifiers 1..507 /organism="Homo sapiens" /db_xref="GDB:466960" /db_xref="taxon:9606" /clone="IMAGE:111343" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia with a modified polylinker; Site_1: Pac I; Site_2: Eco RI 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGTGAAGAATTAATTAAAGACTCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	106 a	125 c	114 g
ORIGIN	125 t	159 t	3 others
Query Match	60.0%;	Score 15;	DB 189; Length 507;
Best Local Similarity	100.0%;	Pred. No. 73;	
Matches	15; Conservative	0; Mismatches	0; Indels
	0; Gaps	0;	
OY	8	cctgccagcatttc	22
Db	192	CATGCCAGCATTTC	178
RESULT	35		
LOCUS	BG467085	514 bp	mRNA
DEFINITION	1A04C06 Bowline Mixed Adipose cDNA library Bos tauros cDNA 5', mRNA		

sequence.
BG467085.1 GI:13396060
EST.
cow.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 514)
Dixon,S.S., Hansen,C., Li,C., Fu,A., Meng,Y., Li,G., Murdoch,G.,
Moore,W. and Christopherson,B.
cDNA's from bovine subcutaneous adipose tissue
unpublished (2001)
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BTNDHDUB (B.taurus CI-B14.5b mRNA for
NADH dehydrogenase (ubiquinone)) in main database at high score of
908.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 514
POLA-No.

FEATURES
source
1..514
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Mixed Adipose cDNA library"
/sex="two males and one female mixed"
/tissue_type="Adipose"
/cell_type="Adipocyte"
/dev_stage="Young adult"
/lab_host="X11-BlueMRF'strain"
/note="Organ: Subcutaneous, Omental, Mesenteric adipose;
Vector: Uni-2APX; Site:1; EcoRI; Site_2: Xho I"

BASE COUNT
115 a 112 c 132 g 155 t

ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 514;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 tgcacgccagcatt 20
|||||
DB 242 tgcacgccagcatt 256

RESULT 36
LOCUS A1957283 524 bp mRNA EST 20-AUG-1999
DEFINITION u185e12.x1 Sugano mouse kidney mRna Mus musculus cDNA clone
IMAGE:2158894 3', mRNA sequence.
ACCESSION A1957283
VERSION A1957283.1 GI:5749992
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
MUSCULUS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 524)
Marrar,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritzer
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterson,R. and Wilson,R.

TITLE	The WashU-NCI Mouse EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: Maria M/WashU-NCI Mouse EST

FEATURES	Location/Qualifiers
source	1. .524

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 1036
High quality sequence stops: 311
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1036 Std Error: 0.00
Seq primer: M13Rpi
High quality sequence stop: 311.
Location/Qualifiers
1..528
source

BASE COUNT ORIGIN	121 a	124 c	110 g	166 t	7 others
----------------------	-------	-------	-------	-------	----------

Query Match 60.0%; Score 15; DB 237; Length 545;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgcca 15
 |||
 DB 505 GCAGGTGATCGCCA 491

RESULT 41
 AM588900/c
 LOCUS
 DEFINITION
 AM588900 547 bp mRNA EST 05-APR-2001
 rat0605.y2 Bird-Rao Meloidogyne incognita J2 meloidogyne incognita
 cDNA 5', similar to TR:055013 055013 BRT3.; contains Alu repetitive
 element.; mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 AM588900.1 GI:7275932
 EST.
 southern root-knot nematode.
 Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima;
 Tylenchoidea; Heterodertiae; Meloidogyninae; Meloidogyne.

REFERENCE
 AUTHORS
 1 (bases 1 to 547)
 McCarter, J., Clifton, S., Chiapelli, B., Page, D., Martin, J., Wyle, T.,
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
 Gibbons, M., Rittler, E., Bennett, J., Franklin, C., Tsagaris, H., R.,
 Rook, L., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
 M., Allen, M., Person, B., Swaller, T., Harvey, N., Schuk, R., S.,
 Shurt, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
 Wilson, R.

TITLE
 JOURNAL
 COMMENT
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 The library was constructed by Uma Rao and David Bird
 (david_bird@u.wustl.edu) at North Carolina State University. DNA
 Sequencing by: Washington University Genome Sequencing Center St.
 Louis.

Seq primer: T3 ET from Amersham
 High quality sequence stop: 519.

FEATURES
 source
 Location/Qualifiers
 1..547

/organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="X10RL"

/note="Vector: ZAP express - PBKCMV (Stratagene); Site 1:
 EcoRI; Site 2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina State University."

BASE COUNT 179 a 76 c 106 g 185 t 1 others

Query Match 60.0%; Score 15; DB 118; Length 547;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatc 22
 |||
 DB 296 CATGCCAGCATTTTC 282

RESULT 42
 A2283519/c
 LOCUS
 DEFINITION
 A2283519 553 bp DNA GSS 27-JUL-2000
 RPCI-23-125E16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-125E16
 , DNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 A2283519.1 GI:9525226
 GSS.
 house mouse.
 Mus musculus

REFERENCE
 AUTHORS
 1 (bases 1 to 553)
 Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akintet,
 B., Levins, M., Megam, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
 and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-125E16.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.bu@tigr.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.bu@tigr.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 125 row: E column: 16

Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..553

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-125E16"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: PBACe3.6; Site 1:
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Size
 selected DNA was cloned into the PBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 135 a 156 c 126 g 136 t

Query Match 60.0%; Score 15; DB 241; Length 553;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgcca 15
 |||
 DB 525 GCAGGTGATCGCCA 511

RESULT 43
 C98142/c
 LOCUS
 DEFINITION
 C98142 567 bp mRNA EST 19-OCT-1998
 C0777.10Z, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 C98142.1 GI:3760888
 EST.
 Oryza sativa subsp. japonica.

ORGANISM *Oryza sativa* subsp. *japonica*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoae; *Oryza*.
REFERENCE 1 (bases 1 to 567)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'

FEATURES
SOURCE Location/Qualifiers
1..567
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C0777.102"
/note="lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid. 1 others
BASE COUNT 190 a 105 c 196 g 75 t
ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 567;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 472 TGCATCGCCAGCATT 458
|||||
6 tgcacgcacgacatt 20
|||||

RESULT 44
AA405993/c 581 bp mRNA EST 09-NOV-1997
LOCUS 2066f09.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742985
DEFINITION 5' similar to contains MER29.b3 MER29 repetitive element ;, mRNA
sequence.
ACCESSION AA405993
VERSION AA405993.1 GI:2063976
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
Kritzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-MCI human EST Project
Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 997 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 498.

FEATURES
SOURCE Location/Qualifiers
1..581
/organism="Homo sapiens"
/db_xref="GDB:5930030"
/db_xref="taxon:9606"
/clone="IMAGE:742985"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGACGACGGACGGCCCAATTTTATTTTATTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 153 a 171 c 131 g 126 t
ORIGIN

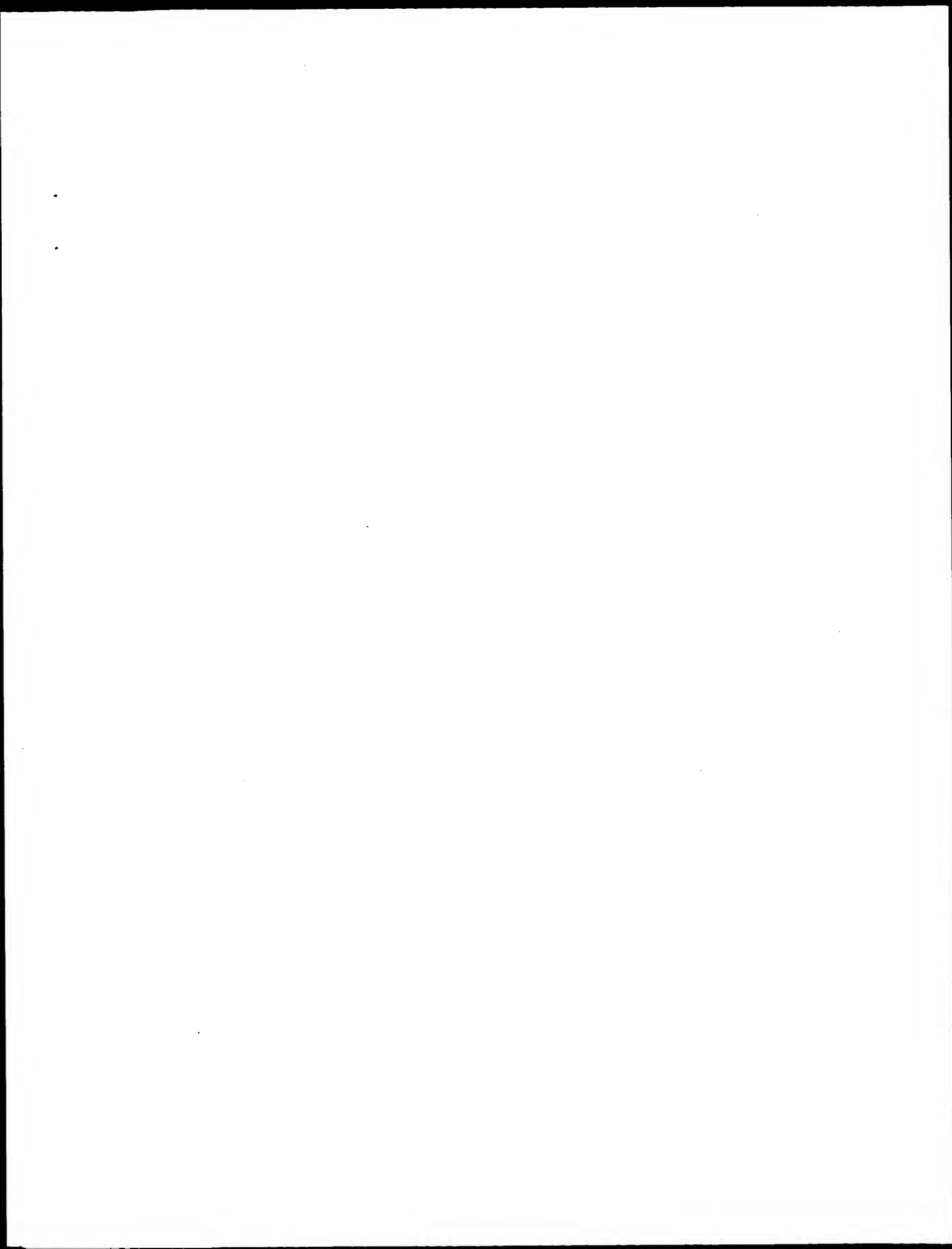
Query Match 60.0%; Score 15; DB 7; Length 581;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 81 CATCGCCACGATTTC 67
|||||
8 catcgccacgatttc 22
|||||

RESULT 45
A0783596/c 605 bp DNA GSS 03-AUG-1999
LOCUS HS_3121_A2_F07_77A CIT Approved Human Genomic Sperm Library D Homo
DEFINITION HS_3121_A2_F07_77A CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3121 Col=14 Row=K, DNA sequence.
ACCESSION A0783596
VERSION A0783596.1 GI:5691150
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 605)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: wallaceu.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3121 row: K column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 605.
Location/Qualifiers
1..605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3121 Col=14 Row=K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in

BASE COUNT E-Coli DH10B"
ORIGIN 187 a 187 c 123 g 92 t 16 others

Query Match 60.0%; Score 15; DB 233; Length 605;
Best Local Similarly 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 gtgcatgccagcat 19
 |||||
Db 248 GTGCATGCCAGCAT 234

Search completed: October 9, 2001, 15:15:32
Job time: 13654 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:46:36 ; Search time 1670.83 seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-7

Perfect score: 25
Sequence: 1 gcatcgccagcattcgatctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 201657

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rtd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_vl3:*
61: gb_vl4:*
62: gb_vl5:*
63: gb_vl6:*
64: gb_vl7:*
65: gb_vl8:*
66: gb_vl9:*
67: gb_vl10:*
68: gb_vl11:*
69: gb_vl12:*
70: gb_vl13:*
71: gb_vl14:*
72: gb_vl15:*
73: gb_vl16:*
74: gb_vl17:*
75: gb_vl18:*
76: gb_vl19:*
77: gb_vl20:*
78: gb_vl21:*
79: gb_vl22:*
80: gb_vl23:*
81: gb_vl24:*
82: gb_vl25:*
83: gb_vl26:*
84: gb_vl27:*
85: gb_vl28:*
86: gb_vl29:*
87: gb_vl30:*
88: gb_vl31:*
89: gb_vl32:*
90: gb_vl33:*
91: gb_vl34:*
92: gb_vl35:*
93: gb_vl36:*
94: gb_vl37:*
95: gb_vl38:*
96: gb_vl39:*
97: gb_vl40:*
98: em_ba3:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 Biob gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7-8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

```

9      25      100.0      5872      9      A93679
10     25      100.0      5872      9      AR101809
11     25      100.0      5872      9      AR101810
12     25      100.0      11022      1      AE000180
13     25      100.0      11501      1      AE005258
14     25      100.0      297816      2      AP002553
15     24      96.0      5526      2      AF250776
16     18      72.0      128      2      EECOB10B
17     17      68.0      479      54      G01446
18     17      68.0      53067      65      AC019940
19     17      68.0      153870      63      AC013431
20     17      68.0      183439      4      AC007809
21     17      68.0      232744      5      AE003705
22     17      68.0      237119      65      AC017740
23     17      68.0      327446      4      AE003500
24     16      64.0      1186      2      AF146614
25     16      64.0      11910      56      AF206717
26     16      64.0      117968      92      HS272E8
27     16      64.0      118100      65      AC017911
28     16      64.0      140825      13      AP002817
29     16      64.0      146081      13      AP001366
30     16      64.0      171705      4      AC007476
31     16      64.0      183937      71      AC037435
32     16      64.0      266133      5      AE003822
33     15      60.0      1389      14      NTPTLG
34     15      60.0      1713      14      NTPECL
35     15      60.0      1907      93      HSSYNTRO5
36     15      60.0      2616      85      AB048915
37     15      60.0      3677      3      PSTAAMH
38     15      60.0      5618      85      AB028980
39     15      60.0      10205      1      AE004320
40     15      60.0      10593      1      AE003860
41     15      60.0      10978      1      AE004795
42     15      60.0      23873      65      AC018333
43     15      60.0      31424      58      AF168792
44     15      60.0      57237      6      CEY62E10A
45     15      60.0      94487      12      AC012394

```

ALIGNMENTS

```

RESULT 1
LOCUS   AR029499      1041 bp      DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION   AR029499.1      GI:5941472
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS  Patton,D.Andrew.
TITLE    Enhanced biotin biosynthesis in plant tissue
JOURNAL  Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
     Source             1..1041
                        /organism="unknown"
BASE COUNT  262 a      273 c      305 g      201 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 gcatcgccagcattcgatcctcgt 25
|||||
Db 90 GCATCGCCAGCAATTCGATCCTCGT 114

```

```

AR034916
LOCUS   AR034916      1041 bp      DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION   AR034916.1      GI:5950521
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS  Patton,D.A.
TITLE    Transgenic plants having increased biotin content
JOURNAL  Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
     Source             1..1041
                        /organism="unknown"
BASE COUNT  262 a      273 c      305 g      201 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcatcgccagcattcgatcctcgt 25
|||||
Db 90 GCATCGCCAGCAATTCGATCCTCGT 114

```

```

RESULT 3
LOCUS   A11530      1084 bp      DNA
DEFINITION Bior gene of E.coli with primers.
ACCESSION A11530
VERSION   A11530.1      GI:490218
KEYWORDS
SOURCE    Escherichia coli.
ORGANISM  Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL  Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
     Source             1..1084
                        /organism="Escherichia coli"
                        /db_xref="taxon:562"
                        /gene="B10B"
                        /gene="B10B"
                        /gene="B10B"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CA00965.1"
                        /db_xref="GI:490219"
                        /db_xref="SWISS-PROT:P12996"
                        /translation="MAHRPRMTLSOYELFEKPLDILFEAQOYHROHDPROYAST
                        LLSITGACPEPDCYKPOSSRYKGLAEARLMEYQVLESARRAKAGSTRPCGAAM
                        KNHPRBMPRIEONVQGVKAMGEACHTLGLTSQAOURLANAGLDYININDTSPEF
                        YGNITITRTYQERLDLTLEKVDADGKIVCSGIVGLGTVKDRAGLLQLANLPTPPS
                        VPINMLVKVGTPLADNDVDADPFIFRTIIVARIIMPTSVYRLSAGREONNEOTQAMC
                        FMAGANSIFYGCKLITTPNPEDEKDLQLPFKRLGINPOQTVAVLAGNDNQQRRLDALMT
                        PDTDEYVNAAL"

```

```

BASE COUNT  271 a      286 c      318 g      209 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcatcgccagcattcgatcctcgt 25
|||||

```

```

Db      113 GCATGCCAGCATTTCGATCCTCGT 137

RESULT 4
E00893  E00893      1121 bp      DNA      PAT      29-SEP-1997
DEFINITION   Genomic DNA encoding biotin Synthetase.
ACCESSION    E00893
VERSION      E00893.1 GI:2169154
KEYWORDS     JP 1986149091-A/1.
SOURCE       Escherichia coli.
ORGANISM     Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE    1 (bases 1 to 1121)
AUTHORS      Hirono, Y., Kojima, T. and Kimura, H.
TITLE        DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
              PRODUCTION OF BIOTIN
JOURNAL      Patent: JP 1986149091-A 1 07-JUL-1986;
              NIPPON SODA CO LTD
              OS Escherichia coli
              PN JP 1986149091-A/1
              PD 07-JUL-1986
              PF 24-DEC-1984 JP 1984272605
              PT HIRONO YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
              C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
              C12R1:19);
              CC strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              CC *source: strain-Escherichia coli Nsl01;
              CC Feature is identified by experimental;
              FH Key Location/Qualifiers
              FT CDS 42..1079
              /product="biotin synthetase".

FEATURES
source      1..1121
             /organism="Escherichia coli"
             /db_xref="taxon:562"
BASE COUNT  289 a 296 c 325 g 211 t
ORIGIN
Query Match      100.0%; Score 25; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gcatgccagcatttcgattcctcgt 25
         |||||||||||||||||||||||||
Db      131 GCATGCCAGCATTTCGATCCTCGT 155

RESULM 5
ECOBIO  5793 bp      DNA      BCT      28-FEB-1994
LOCUS    E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION   (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
              protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION    J04423
VERSION      J04423.1 GI:145422
KEYWORDS     7,8-diamino-pelargonic acid aminotransferase;
              7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
              bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
              synthetase;
              Escherichia coli (strain K-12) DNA.
SOURCE       Escherichia coli
ORGANISM     Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE    1 (bases 1 to 5793)
AUTHORS      Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and
              Johnson, O.

TITLE      The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL    J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE    89066784
COMMENT    Draft entry and computer-readable sequence [1] kindly submitted by
              A.Otsuka, 09-NOV-1988.

FEATURES
source      1..5793
             /organism="Escherichia coli"
             /strain="K-12"
             /db_xref="taxon:562"
             /complement(98..574)
             /note="ORF 1"
             /codon_start=1
             /transl_table=11
             /protein_id="AAA23513.1"
             /db_xref="GI:451568"
             /translation="MKLISNDLRDGLPHRRHVNMGMDGSDNISPILAMDDVPAQTK
              SEVVVCYDPDAPGTSGMMHVVVNLPAQTRVLPQGFSGGLVAMPDGLQTRPDGKTG
              YDGAAPRGKGTHERYIFVHALDIERIDVDGASAMVGFVHFSLASATAMFS"
             /complement(633..1925)
             /gene="bioA"
             /complement(633..1925)
             /gene="bioA"
             /codon_start=1
             /transl_table=11
             /product="7,8-diamino-pelargonic acid aminotransferase"
             /protein_id="AAA23514.1"
             /db_xref="GI:457106"
             /translation="MTTDLAPDQRHIMWPTTSMTPSPVPVVSABECELLSDGR
              LVDGSMWMAAIIHGYNHPQLNMAKSDIDAMSHVWGGITNAPALELCKRLVAMSGRN
              ALVEGVADSGSVAVEAVAMKMAILOYWOKGEARQPLTFRNQYHGDYFGASVDPDN
              SMHSIKMGYLPNLPAFAPASRMDGEMDMDVGFARLMAHREHIAVILEPIYOGA
              GCMRWHPMLKIRIKICDRGILLIADIEITGPGTFCACBAANASTALLESQDMO
              LTGGTWTLSATITTRVAVETISNGFAGCPHMPGPNPLCAANASTALLESQDMO
              QGVADIEVQLREQLAPARDAEWADVRLGATGVETTHPVMAALQKFPVQGVWIR
              PGKLIYIMPPYIILPQQLRLTAANVNAVODETEFCO"
             /2012..3052
             /gene="bioB"
             /2012..3052
             /gene="bioB"
             /codon_start=1
             /transl_table=11
             /product="biotin synthetase"
             /protein_id="AAA23515.1"
             /db_xref="GI:145425"
             /translation="MAHRRWTLISQVTELEFEKPLDILFEAOQVHRHPPROQVST
              LLSIKTGACPEDECKTPQTSRYKTGIEARLMEVEQVYLESARKARAAGSTRGCMGAAM
              KNPHERMDMPYLEDQWGVGRAMGLEACMTLGTLSQSOQLRANAGLDIYNHNDTPER
              YGNIITRTYQERLDTLEKVRDAGIKVCSGGIVGLGTEVKRAGLLDLQLANLPTPES
              VPINMLVKVGFPLADNDVDADFRTIYAVARIIMKPTSVYRLSAGREOMDQOAMC
              PMAGANSIFYGCKLFTPNBEDKDLQLEFRKIGLNPQYPAVLADNEQOORLEQALMT
              PTDREYNNAAAL"
             /3049..4203
             /gene="bioF"
             /3049..4203
             /gene="bioF"
             /codon_start=1
             /transl_table=11
             /product="7-keto-8-amino pelargonic acid synthetase"
             /protein_id="AAA23516.1"
             /db_xref="GI:145426"
             /translation="MSWQEKINNALDARRAADAARRRYPAVGAGRWLVADROYLNF
              SMDYVGLSHHPQIIRAMOGAGPFGIGSGSGSHVSGSYVHQALEEELAEWLGYSRA
              LPTISGPAANOAVIATAMMAKEDRIADRLSHSLPAAISLPSQLRFRFANVDYTHLAR
              LLSRPGCGQGVYVTEGVESMDGSAPLAEIQOYTOOHNMIMVDDHNRFGHVGISEQGRG
              SCHLQKVEPELVVTTGKRGVSGAVALCSSYTAADYLLOPAHLLYSTSPAPAQAL
              RASLAVIRSDGDAARRKALALITFRAGVQDLPFTLABSCSAIOPLIYGDNSRAQL
              AEKLRQGCWVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLIHNG"
             /4190..4945
             /gene="bioc"
             /4190..4945
             /gene="bioc"
             /gene="bioc"

```

```

/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVANKOAIAAFGRAAAHYEHOADLOROSADALLAMLPORRYT
HVLADGCGPGMWSRHREHRAOVTALDISPMLVOAKODADHYLAGDIESPLATA
TFDLAWSNLAVQWCGNSTALRELRYRVRPGVYAFPTTIYOGSIPEHQAQAVDEBP
HANRFLPDEIEOSLNGVYHIOHTOPTLMDLDSMRSLKIGATVHLHGDRPLL
TRSLORLQRLAMPQOQGRYPLTYHFLGVARE"
4938..5597
/gene="B10D"
/feature="B10D"
4938..5597
/gene="B10D"
/codon_start=1
/transl_table=11
/product="gethidiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYEVTGDTFVGKTVASCALLQAKRAGYPRAGYKPVASGS
EKPEGRNSDALALORNSLIDYATVPNTFAEPISPHIISAQEGRIESTLVASG
LRALEQADWYIVGAGCGMPTLSDPTFPADWTQDOLPVLVGVKGLGCTINAMLTA
OYIOHAGLITLAAWYANDVTPPGKHAEYMTTLTMTIPRCWERSPLQKIQKROPESS
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcatgcagcattcgcctcgt 25
Db      2101 GCATGCCAGCATTCGATCCTCGT 2125

RESULT      6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1 GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLES      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 1 14-APR-1994;
COMMENT
      LONZA AG (CH)
      Other publication PL 308301 950724
      Other publication CA 2145400 940414
      Other publication AU 4820293 940426
      Other publication HU 71781 960228
      Other publication SK 42095 951108
      Other publication CZ 9500809 950913
      Other publication FI 951547 950331
      Other publication JP 8501694T 960227.
      Location/Qualifiers
        1..5872
          /organism="Escherichia coli"
          /strain="DSM498"
          /db_xref="taxon:562"
          /clone="PB030A-15/9"
          1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
              /standard_name="PROMOTER PTAC"
              /evidence=experimental

```

```

RBS      105..119
/standard_name="B10B RBS NO.9"
/evidence=experimental
117..1157
/gene="B10B"
117..1157
/gene="B10B"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRRMTLSQVTELEKPLDILLEAQOYVHRHEDPROVOYST
LSTIKGACPECKCYCPOSSRYKGTGLAEERLMEVROYESARAKAAGSTRFCGAAN
KNPHERDMPLYEOMVOGYKAMGLEACMLGTLSQAOARLANAGDITYNHLDPSPER
VGNITTRTYORLDTLEKVRAGIKVSGGIVGIGTVKORAGILLQLANLPPTES
VTINMLVKKGTPLDNDVDFTRTIYARIMMPTSYVRLSGRQNMEOQAMC
PMAGANSLEYGKLLTTPNEBDKDLQLEFKILGINPOQTAVLAGDNEQORLEQALMT
PDTDEYNAAL"
2284..2297
RBS
/standard_name="B10C RBS"
2295..3050
/gene="B10C"
2295..3050
/gene="B10C"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVANKOAIAAFGRAAAHYEHOADLOROSADALLAMLPORRYT
HVLADGCGPGMWSRHREHRAOVTALDISPMLVOAKODADHYLAGDIESPLATA
TFDLAWSNLAVQWCGNSTALRELRYRVRPGVYAFPTTIYOGSIPEHQAQAVDEBP
HANRFLPDEIEOSLNGVYHIOHTOPTLMDLDSMRSLKIGATVHLHGDRPLL
TRSLORLQRLAMPQOQGRYPLTYHFLGVARE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/gene="B10A"
3750..5039
/gene="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDDLAPDORHIMHTPMTSPYPPVVASGECETLISDGR
LEVDGSSMMAIHGYNHPOLNAAMKSQIDMASVYMGCTHAHAIELCRTVMTPOP
MHSILMKGYLPENLPAPQSRMDGEMDERMVFARLMAAHRHEIAVITEPTVOAGS
GMRYHPEMLKRIKICDREGILLIADEIATGFGRTGKLPACHEAIEADITLIGKAL
TGGTMTLSATLTREVAETISNGEAGCFMHPFEMNPLACAAANSIALILLEGDMOO
QVADIEVOLREQLAPADAEVAVDVIAGIVETTHPPVMAALQKFEVGGVWLIRP
EGRLIYMPYIILPQDLQRLTAAVNNAVODEIFFCO"
5088..5100
RBS
/standard_name="ORF1 RBS"
5096..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGDKILPHRVFNGMGVGDGNISPHLAWDDVPAGTK
STVVTCTYDDAPFTGSGMHWVYNLPADTRVLPGCFGSLVAMPDGVQOTRPFCKTG
YDGAAPKGETHRIFVYHALDIERIDVEGASGAMVGFNHFHSLASASTTAMS"
stem_loop
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gcatcgccagcatttcgactcgcgt 25
      |||||||||||||||||||
Db      206 GCATCGCCAGCATTTCGATCCTCGT 230

RESULT 7
LOCUS      A38251      5872 bp      DNA      PAT      05-MAR-1997
DEFINITION      Sequence 6 from Patent WO9408023.
ACCESSION      A38251
VERSION      A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 6 14-APR-1994;
      LONZA AG (CH)
COMMENT      Other publication PL 308301 950724
      Other publication CA 2145400 940414
      Other publication AU 4820293 940426
      Other publication HU 71781 960228
      Other publication SK 42095 951108
      Other publication CZ 9500809 950913
      Other publication FI 951547 950331
      Other publication JP 8501694* 960227.
      location/Qualifiers
FEATURES
Source      1..5872
      /organism="Escherichia coli"
      /strain="DSM498"
      /db_xref="taxon:562"
      /clone="PBO30A15-9"
      1141..1156
      /standard_name="BIOF RBS"
      1154..2308
      /gene="BIOF"
      1154..2308
      /gene="BIOF"
      /standard_name="8-AMINO-7-OXONONAMATE SYNTHASE"
      /EC_number="2.3.1.47"
      /codon_start=1
      /transl_table=11
      /number=2
      /evidence=experimental
      /product="KAPA SYNTHASE"
      /protein_id="CAA02329.1"
      /db_xref="GI:2294850"
      /translation="MSMOEKINALLDARRAADALRRRYPAAGAGRMLVADDDRYLNF
      SSNDYLGISHHPDIIIRAMQOGEAFQIGSGSGVSGSVYHQALEELBWLSTSA
      LLPTISGFANQAVIAAMAKEDRIADRLSHASLLEASLSPQLRRFAHNDVYHLAR
      LLSPCPCQGMVTEGVFSGMDGSAPIAETIOVYIOHNGWIMVDAGTGVIGBGRG

```

```

SCWLQKVKPELLVYTFGKGFVSGAIVICSTYADYLLQFARHLITYSMPPAQAL
RASIAYIRSDGDKRREKLAALITRRFAGYODLPETLADSCSAIOPLIYDGNRALQL
AEKLRQGCWTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEVTGDTFEVGTVAACALLOAKAGRYTAGYKPYASGS
EKTPGCLRNSDALIQRNSSIOLDYATVPYTPAEPSPHISAEGRPIESLYMSAG
LRALFOQADWYIVBEGAGMPPRLSDTTFEADWYQEOPLVILYVVGKGINHAMLA
QVIOHAGLTLGWAVANDVTPPGKRAAEITMTLTMTIRAPLLGELPMLAENENATGK
YINAEVDSITGFTSRL"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gcatcgccagcatttcgactcgcgt 25
      |||||||||||||||||||
Db      206 GCATCGCCAGCATTTCGATCCTCGT 230

RESULT 8
LOCUS      A93674      5872 bp      DNA      PAT      22-JAN-2000
DEFINITION      Sequence 1 from Patent EP0798384.
ACCESSION      A93674
VERSION      A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 1 01-OCT-1997;
      LONZA AG (CH)
COMMENT      location/Qualifiers
FEATURES
Source      1..5872
      /organism="Escherichia coli"
      /strain="DSM498"
      /db_xref="taxon:562"
      /clone="PBO30A-15/9"
      1..96
      /function="PROMOTER PTAC"
      /evidence=experimental
      23..48
      /standard_name="PROMOTER PTAC"
      45..49
      /evidence=experimental
      105..119
      /standard_name="BIOB RBS NO.9"
      /evidence=experimental
      117..1157
      /gene="BIOB"
      117..1157
      /gene="BIOB"
      /codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPMTLSOVTELEFPEKPLDLLFEAQOVHROHDPVOVST
LSTKTCACPEDCKYCPOSSRYKIGLFAERLMEVEVLESARKAKAGSTRPCMGAW
KNEHEDMPYLEQWGVGKAMGLEACMTIGTLESQAORLANAGLDLYNNILDTSEF
KGNITRTYQERLDLEKVDAGIKVCSGGIVGLGTYKAGLLOLANTPTPES
VPIMLVKVGTPLADNDVDADFIRITIAVARIMPTSVYRSLAGREOMNEOTQAC
PMAGANSIFPGCKLTTPNEEDKDLQLFKLGILNPQGTAVLAGDNBQOORLEQALMT
PDIDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVVKQATAAAGRAAHYRQHADLQOSADALLAMPORKYT
HVDAGCGPGMSRHRHRRRAOYTALDLSPPMLVORQKDAADHLAGDIESLPLATA
TPDLANSNAVOCGNLSTALRELYRVVRPKGVAVFTLVGSLPELHQAQVADERP
HANRELPPEDEIOLNGVHVOHHIOPITTFEDALSAMRSLKGIGATHLHEGRDRI
TRSOQLQLOLAMPQOQGRYPPLYHFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADFDRHIMHPTSMTPLEVPVVSABGCELLSDGR
LYDGMSSMAAITHGYNHNPOLNAKMSODAMSHVMGGITTHAPATELRKIVAMPOR
LECVLADSGVAVEAMKALQWQAKGEARORLITRNGHGTPTGMSVCDPDS
MSIMKGYLPENLAPAPOSRMDGEMDERDMGPRLMARHHEITAAVITEIVGAG
GKRYHPEVLEKRIKICDEGLILTADEATGFGTKLFCENAEIADILCLKAL
TGGMTLSATLTREVAETISNGEGCEMHGTFMGNPLACAAANASLAIIESGDMOC
QVADIEVOLREQLARDAEMVADVIVGALGVETTPVNNALQKFFVEGVWIRP
FGKLIYMPPIYILPOOLRLTAANRAVQDETFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLPRHRVFNMGVGDGNTSPHLAMDVYAGTK
SEVVTCYDPAVPGSGMWVYVNPADTFLPQCGSGIYVHSDVLQITTTGCKG
YDGAAPKRGEMHRYFTYHALDIERIDVDEGASGAMGFINVHSHLASASTAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a, 1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  gcatgccacattgcattccgt 25
Db      206 GCATGCCACACATTGCATCTCTGT 230

RESULT      9
A93679      A93679      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 6 from Patent EP0798384.
DEFINITION  A93679
ACCESSION  A93679
VERSION    A93679.1  GI:6741867
KEYWORDS   .
SOURCE     Escherichia coli.
           Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE   1 (bases 1 to 5872)
            Birch, O. and Brass, J.
            Biotechnological method of producing biotin
            Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES
     source          location/qualifiers
     RBS             1..5872
     gene            /organism="Escherichia coli"
     CDS             /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PB030A15-9"
                     1141..1156
                     /standard_name="BIOF RBS"
                     1154..2308
                     /gene="BIOF"
                     1154..2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINALLDARRAADALRRRYPAAGAGRMIVADPOYLNF
                     SSNDYGLSHPOIIRAMQGAOPGIGSGSGSHVGYVHQALEELAEMLGYSRA
                     LPTSGAANDAVIYAAMAKEDITADRLSHASLSRSTPSQLRRFANDVTILAR
                     LLASPCGQOQVTEGVFSMDGSDAPLAEIQVTOQHNGLWMDHAGVIGEOGRG
                     SCHLQKVPPELLVVTGFGVSGAVALSSYVADYLDFARHLIYSTMPAPQAOAL
                     RASLAVIRSDGDAKREKLAALITRRFAGVODLPFLADSCSAIOPLIQDMSRALOL
                     AEKIROGQCWVTAIRPTVPAGTARLRLLTAHHEMODIDRLLEVLHGNG"
                     3030..3045
                     /standard_name="BIOB RBS"
                     3043..3753
                     /gene="BIOB"
                     3043..3753
                     /gene="BIOB"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVYGTQTEVNGKTVASCALLQAAAGCFRTAGYRPVPSGS
                     EKPEGLRNSDALALORNSSLQIDATVAVPTTFAEPTSPHITSABGRITESLVASG
                     LRLEQADWNLVEGAGWFTPLSDFTTADWTOEQPLVITLVGKLGCTINHAMLTA
                     QVILHAGTLTAGWVANDVTPPKRRRAEYMTLLTRMIPALDGEIPMLAENPNAAATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN YINLAFVASTLIGFTSRL"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccgcgt 25
 |||||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 10
 ARI01809
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing diotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccgcgt 25
 |||||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 11
 ARI01810
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing diotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN /organism="unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccgcgt 25
 |||||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 12
 AE000180

LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 000096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
 Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
 Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
 Mau,B. and Shao,Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: markborov@gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (http://cgsc.biology.yale.edu). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source

Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

repeat_region
 <1..130
 /db_xref="taxon:83333"
 /note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
 complement(147..1430)
 gene
 /gene="ybhc"
 /note="b0772"
 complement(147..1430)
 CDS
 /gene="ybhc"
 /function="orf: Not classified"
 /note="f427; 98 pct identical to fragment YBHC_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVRLALALAFGVTLTACSSPPDQRPSPDQAPGTSRRP1
 LSAAQNPDAOHYFASLTGGAAMNPSPITLPADPVPVGPAGTQGVHTTIOAAD
 AAIKRTNKROYIAYMDEYQGVYVPAADGILTYGEPIDVKIGSLDGMSPA
 DMHVDVNPCKYMGKPAWYMGSCOSKRSDSIGLCSAVFMSQNNGLQNLNTT
 LGSVDAGNHPAVALRTDGDVOYIINNVIILGRNTFVNSGVONLEINROPRITVT
 NSYLEGVIVSGGAVNPDPTEPRVNSRTOEAYVEAPATISNITGYGLAANSRN
 AFSDGVAQIGKSLVDANTNGQVYTRDSAINESGNTAKPWADAIVISNRPAGNTGSD
 DNDEIQRNLMDNTNRMWEYNNGVSKVAEAKK"
 complement(1507..1535)
 /note="factor sigma70; predicted +1 start at 806574"
 complement(1582..2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582..2058)
 /gene="ybhB"
 /function="orf: Unknown"
 /note="f158; 99 pct identical to YBH_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNLRDGDKLPRHVGNGXGDDNI SPHLAMDVPAGTK
 SFVVCYDPAPGSGMMHWVYNLPADFRVLQGGSGGLVAMPDGLQTRIDFGGTG
 YDGAAPKGETHRIYFVHALDIERIDVDGASGAMGVFVHPSIASITAFMS"
 complement(2108..2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rnas predicted site"
 complement(2117..3406)
 /gene="bioA"
 /note="b0774"
 complement(2117..3406)
 /gene="bioA"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers: Biotin"
 /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDDIAFDQRHIMHYTSMTEPLVYVYVSAEGCELLISDGR
 LVGMSWMAIHGVNHPQLNAAMKSOIDMSVMVFGTTHAIELCRKLVAWTPQP
 LECVFLADSGSAVEVAKMALQYQAKGARGORFLFRNGYHGFDEFGAMSCDDPNS
 MHSIMKGYLPENFLFAPQSRMDGEMDERMVFARLMAAHRHIEIAVIEPVOGAG
 GMRVYHEPMLKRIKICDREGTLLADEITGCGPGKLEACEHAETAPDIICIGAL
 TGGMTLSATLTREVAETISNCEAGCGMPEMKPLACAAANSIAPIITGSGMOO
 OVADIEVQLREQLAPADAEVADYRVLGAGVETTHPVNMAALOKFEVEGCVWIRP
 FGKLTLPPIIILPQQLRQLTAAVNAVDETFEQ"
 complement(2193..2221)
 /gene="bioA"
 /note="factor sigma70; predicted +1 start at 807260"
 3411..3450
 /note="central position to bioB promoter: -20"

protein_bind
 /bound_moiety="BioB documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic evidence for the site"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="BioB documented site"
 3413..3441
 /note="factor sigma70; promoter bioB; documented +1 at 808525"
 complement(3447..3473)
 /note="factor sigma70; promoter bioA; documented +1 at 808515"
 3493..4533
 /gene="bioB"
 /note="b0775"
 3493..4533
 /gene="bioB"
 /EC_number="2.8.1.-"
 /function="enzyme: Biosynthesis of cofactors, carriers: Biotin"
 /note="f346; 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRKWTLSQVTELEFKPLDLLFPAQYVHROHFDPROQVST
 LLSIKTGACPEDEKCYQSSRYKTGLEERLEMEYQVLESARKKAAGSTFECAAM
 KNHREMDPELEOMVOGVKAMGLEACMTLGLISEQORLANAGIDYVNNHLDPSER
 YONIITTRTYGERLDTLEKVRDAGIKVSGIVIGAGTVKRDAGILLQALNLPPEP
 VPIMLVYKVGKPLADNDVDPADEFRTIYAVARIIMPYSYVLSAGREONFQOAMC
 FMAGANSYFGCKLITTPNPEEDKQQLFRRLGLNPQOTVVLADNEDQQLBEALMT
 PPTDEYNNAL"
 4530..3684
 /gene="bioF"
 /note="b0776"
 4530..3684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers: Biotin"
 /note="f384; 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINMALDARRADALRRRYPVAGAGRMLVADROYLNF
 SSNDYLGSHHPQIIRAWQGAQEDFGTOSGSGHVSQSVYHQALEEELAWLSTRA
 LLFISGFANQAVIAAAMAKEDRIAADRLASLSLEASLSPSQLRRFAHADVHLAR
 LLASPCGOQWVTEGVFSGMDGSAPIAEIQOYVQOHNHGMVLDADHGTGVTGCGGK
 SCMLQKVPBELLVTFGKGFGVSGAVALCSSTVADYLLQFARHLITYSTMPAPAOAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcacgacattcgatccctcg 25
 |||||
 Db 3582 GCATGCCGACGATTCGATCCTCGT 3606
 |||||

RESULT 13
 AE005258
 LOCUS 13501 bp DNA BCT 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155
 ACCESSION AE005258 AE005174


```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
              Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
              Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
              Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
              Welch, R.A. and Blattner, F.R.
TITLE        Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL      Nature 409 (6819), 529-533 (2001)
MEDLINE      21074935
PUBMED       11206551
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
              Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
              Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
              Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
              Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
              Welch, R.A. and Blattner, F.R.
TITLE        Direct Submission
JOURNAL      Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source       1. 13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohaemorrhagic"
              <1. 7576
              /note="O-island #36; Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; Cyclic prophage
              CP-933K; Includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66. 665
              /gene="lomK"
              /note="20981"
              66. 665
              /gene="lomK"
              /function="putative membrane; other or unknown (Phage or
              prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              gi175327891|gb|AAAF63231.1|AF151091_2 (AF151091) lom
              [prophage P-B1bA]"
              /codon_start=1
              /transl_table=11
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AA055137.1"
              /db_xref="GI:12513752"
              /translation="MRKVCALILSAICLANSGVPANASEHSTLSACYLHASTDAG
              SDDLNGINVKRYEFTDTLGLITSPFSANADECKTHSDTRMHEIDYRNKWSVMAG
              PSVRNEMFSAVAMAGVAVSRVSTFGDPRVTNKRTHDVLTLGSDDARYSNISLW
              GAGVOFNFTESAVADVAVEXSGSDMRTDGFIVGVYKF"
              719. 2041
              /gene="20982"
              719. 2041
              /gene="20982"
              /function="putative structure; Structural component (Phage
              or prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from GenPept 118 :
              gi145854361|gb|AAD25464.1|AF125520.59 (AF125520) putative
              tail fiber protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="putative tail component of prophage CP-933K"

gene
CDS
        /protein_id="AA055138.1"
        /db_xref="GI:12513753"
        /translation="MGMAAYOISGVLRKDGAKGPIONCTIOLKARRNSTXYVNTVASE
        NPDEGRISMDYEQYQSVTLTVLVBGFPPSHAGCTISYRDSOPGTLINFLGMPEDDAR
        PEALRFQDMVEEARHAEAKRNAGEETSRKNGISASAEBSANPNIDMSADSDAR
        SAROAESAASAKKSEASSSASAEAAQAKAESIQSADALSKRTAESAGNAROA
        TSTSEKAREASQAQSAQSRIFAEDAVNRIPTVVPGPPGEGEPAPGPKDGK
        RBDTPGATGREGPGDGTGPAGPGPGGDRGEGTGLTNAGPGPGPGGGAAGA
        PGPGKPGKGAAPYGATGPGPGPKDDEPTOIREFGLMRILETNSYGMFPGTGALI
        TGLTFLDRDPAQVOVGMPQHLQVRRGDPQMDQVGLDVGSDTGTGTC"
        2004. 2312
        /gene="20984"
        2004. 2312
        /gene="20984"
        /function="orf; other or unknown (Phage or Prophage
        Related)"
        /note="Residues 1 to 102 of 102 are 98.03 pct identical to
        residues 28 to 129 of 129 from GenPept 118 :
        gi145854371|gb|AAD25465.1|AF125520.60 (AF125520)
        hypothetical protein [Bacteriophage 933W]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AA055139.1"
        /db_xref="GI:12513754"
        /translation="MRKAVIOAELENDMNLRLKMQSLGCGSKHDCENRSLTAQLR
        LCPADLLESDENCTIPEODRVTTOGVVILLADKKQIQCVVRPLQLLRADGTWENIGMK"
        2489. 3469
        /gene="20985"
        2489. 3469
        /gene="20985"
        /function="orf; other or unknown (Phage or Prophage
        Related)"
        /note="Residues 13 to 326 of 326 are 55.73 pct identical
        to residues 18 to 331 of 336 from GenPept 118 :
        gi169603671|gb|AAAF3527.1 (AF170176) hypothetical protein
        predicted by Glimmer (Salmonella typhimurium LT2)"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AA055140.1"
        /db_xref="GI:12513755"
        /translation="MLSPRTTFPHNSVNIYVSSPCQVSPAGKEVELKVIDEPTILF
        OMFEPPERYKKDEVPITVNMKQPIYDNTNARISSDMIGIFVGDSPVAKTAFS
        KLERDENVNIITREVDPSMVKDRIKSDYHDIICRQRLRTEDDKRBYTLNLEELR
        ELISKADDSLIMTAKRNAMDFEFNALLAKAGELFRCTYNTKNGISIFGSGCIYLD
        MDMLILGKLGCTIYAPDGISMHWDRNDSVNIETSAITVNRNSIPALBELSGFMSKVD
        AHVYDGLGKGVAKKYFNPPLNHNHCFIEFNHPIINMNSOYTCSSW"
        3530. 4522
        /gene="20986"
        3530. 4522
        /gene="20986"
        /function="orf; other or unknown (Phage or Prophage
        Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AA055141.1"
        /db_xref="GI:12513756"
        /translation="MKIPSLDSNMFNSAPAGYSAPITAPNRENAVADYVLDIGKRIPL
        SAADLSNYESVIRAVHDSRSLIDQHTVDMIGNTVLDAISRSQTFRDVAVGIRINKE
        VHTGCKRYRNEYELNESSYKIDIDIOSLTJCNELYEDVGCEPIFPLCEAGENDEPY
        VTSVAPDRTPSYEMSGELIHHTHHTGNSDGSQDSIDIEIGPELILARVAOELG
        WSPDFKGVAPEREHRLRLRNALRQAMRNRENEPRAFEELGTISRYEASPPFT
        EYSAVSNIGYGFIOQHDPELAINDNLDANQIOLTHGAPYITFGDYVKNHQ"
        5350. 6231
        /gene="20989"
        5350. 6231
        /gene="20989"
        /function="orf; other or unknown (Phage or Prophage
        Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513757"
/translation="MLSFVSNLCCSNMSTLRTITSPDNRLSSRDAAVHSDNGAV
KGVNTRYVATDNKFCVTRSHSGFTNLHLRLGPKGEISRKIEVNLNAPSAM
ERGIHNSRDPDLPVDAPELPSVDNRLSPGNVIGKGNNAVYEDADATKVM
FTTSQSEVTEVRCFNQYAGSAKRIYVNDGIDIGIMDKINSESLNISIPAO
AEHATYDMFDEQKGLFVDTETNVLVYRAKNEFNPIDISINSDRWSQSQIMO
SYHGKODLSVYLSKI"
6462..7160
/gene="Z0990"
6462..7160
/gene="Z0990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLIVLHOSRSSMSDPIESLVASSYQMTKNPQLRF
OGDHNHYQDIEALDKTGSTETGRVLLNATESISRLKSETVVIHNSSLGVMAHD
IDAEHNGTGSDFHCNLAAYPCGEGISVADHATVPHLHLVPHNNGERLAKES
SRASQKYSPLLEAETVGLGAFSEVLSSEKNEFHEIGMPTRTSPXDSALTIDONT
VSLGFQOVRLLPPL"
complement(7668..8144)
/gene="ybhB"
/note="Z0992"
complement(7668..8144)
/gene="ybhB"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: 80773"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPLRHRYFNGMGYDGNISPHLAMDVPAGTK
SEVYCYDPAPTGSGMMHVVYVNLPAIDRVLPQSGSLVAMPQGVLOTRRDCKTG
YDCAAPKCEHRYITVHALDOVERIDVDEGASGAVNGVNHFSLASASTAMPs"
complement(8203..9492)
/gene="BioA"
/note="Z0993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best local Similarity 100.0%; Pred. No. 0.00019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgcgcagcattcgatcgcgt 25
|||||
Db 9668 GCATCGCGACGATTCGATCTCGT 9692

```

```

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1. (sites)

```

```

AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, T., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
TITLE
Complete nucleotide sequence of the prophage  $\phi$ 2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
TITLE
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
TITLE
Complete nucleotide sequence of the prophage  $\phi$ 1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsuno, E., Nakayama, K., Murata, T.,
Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
TITLE
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
TITLE
Direct Submission
JOURNAL
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
COMMENT
Location/Qualifiers
SOURCE
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEYQAKOLFAPRGILPAPGVYACTTPPEAREASKTGAGPV
VKCOVHAGGRGKAGGVVYVNSKEDIRAPAEHNLGKRLVYQDPAQPNQIIVKAT
DIKELTIGAVYDRSSRRVFMASSTGGEIVKAEETPLTHKVALDPLTGPYOG
RELATFKLGEGLKLVQDFTKTEGLATIFLERDLALIEIPLVITKQGLDICDGLIGA
DGNALFRQDPLREMRDQSDPREAQAQWELNVALDNGICVMVAGACILAMGTMDIV
KLHGEPANFLDVGGATKERYEAFKIIISDPRKAVLVNIFGCGIVRCDLADGIG
AAVEGVNVVVRLEGNNAEIGAKKLADSGNIIIAKGLTDAQGVAAVVEGK"

```

```

gene
1245. .2114
/gene="ECS0754"
CDS
1245. .2114
/gene="ECS0754"
/note="similar to SUCD_ECOLI g1178694 percent identity
100 in 289 aa (conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/transl_table=11
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/transl_table=11
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/transl_table=11
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/transl_table=11
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/transl_table=11
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcgcacattcgatctcgt 25
|||||
Db 92400 GCATGCCACGATTTTCATCTCTCT 92424
|||||

RESULT 15
AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION
Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
modC-biA intergenic region, DNA-aminotransferase BioC (bioc),
biotin synthase BioB (biob), GABA-aminotransferase BioF (biof), and
biotin biosynthesis protein BioC (bioc) genes, complete cds; and
dethiobiotin synthetase BioD (biob) gene, partial cds.
AF250776
ACCESSION AF250776.1 GI:12620124
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W. R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196

PUBMED 1113432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streil, W. R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..5526
 Location/Qualifiers
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGRKLPFRHVFNGMGYDGNISPHLAMDVPAGTK
 SEVVTCTDEPAETGCGMHVNVNLPADRVLPQGGSLVAMPDGLQTRDPFGAG
 YDGAAPPKGETHYIFTVAHLDERIDVDGASGAMVGFVHFHSLASITAMSS"
 complement(587..1876)
 /gene="bioA"
 /complement(587..1876)
 /gene="bioA"
 /note="7.8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAFDQRIHHPYVSMTPSPVPSAEGCELLSDGR
 LVDMSSMAALHGVNHPQLNAAMKSQIDAMSHVFGCTTHAPATELCRKLVAMPQ
 LKCVFLADGSAVEAYAMKALOYMOAKGEARQPLTFPFGHGFPGMSVCDGDS
 MSLMGVGLPENLFAPOSRMDGEMDEDMGFARLMAHHEHIAAVITEIYOGAG
 GRMYHPDWLKRIRKIDREGILLTADETATGFGGRKLFACELAEFIDILCKAL
 TGGIMTLSATLTREVAETISNGEAGCFMGGTPMGNPLACAAANASLAIIESGMWO
 OVADIEVQLRDLARADENAVDRVLCATGVETTHFVNNAALQKFFVEQGVIRP
 EKLITLMPPTLITLQDLQRLTAAVNAVADDETFEQ"
 1963..3003
 /gene="bioB"
 1963..3003
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPMTLSQVTEFEKPLDILFEAQQVHRHOFDDROYVST
 LLSITGACPEDCYKCPQSSRYKTSLEAEPLMEVQYLESARKAKAASSTPFCGAM
 KNPENRMDPYLEOMQYKDGCLGAEPLTGLTSBSQARLANAGLIDYINHLDTSPER
 YGNITTTTRYQERLDTLKVADGAKIVCSGGIVGIVGVKRAGLILLOLANIPPPES
 VPIANLVKGTPLADNDVDAFDFIRITAAARLIMPSTSYVLSNGRQNMEOQAMC
 FMAGANSLEFGCKLLTTPNPEBKDLQLEFRKLGNPQQTAVLAGDNEQOQRLQALMT
 PPTDEYVNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /gene="bioF"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWOKTINALDARADALRRRYPAVAGARWLVDADROYLNF
 SSNDLGLSHHPQITIRAMQGAQEQFGTSGSGVSVVHQALFEELAEMLGYSRA
 LFTISGFAANOAVIAAMAKEDRIADRLSHASILEASLSPQLRFRANNDVYHLAR
 LLASPCGOQLVTEGVSMQDSAPLAEIQVITQGNHMLMVDANGTGVICGQNGC

gene
 CDS
 4141..4896
 /gene="bioC"
 4141..4896
 /gene="bioC"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein BioC"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MAIVNKOATAAFGRRAALYEOHADLQROSDAVLLAMPKRYT
 HYLDAGCGPQWMTTRHREHRAQVTLALDISPMLVOAROKDAADHYLAGDIEPLATA
 TFDLWMSNLAVQWGNLSTALRELYRVRSYVAFTLVQGSLLPELHOAOAVDERP
 HANRFLPPEDEILOSNGVYOHIOPTLIMFDDALSMRSLKIGATIHHEGRDRLT
 TRSOLQRIOLAMPQOQGRPYTLTHFLGVARE"
 4883..>5526
 /gene="bioD"
 4883..>5526
 /gene="bioD"
 /note="PTB synthetase"
 /codon_start=1
 /transl_table=11
 /product="delnibiotin synthetase BioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYFVGTDTENGKTVASCALLQAKAAGRTAGYKPYAS
 GSEKTEGLRNSDALIQRNSLQLODPAVAPYFAPRTPHITSQEGPDISSVMS
 SGLRLEQDQDWIVVEGACGCFPTLSSTFFPADVYTOEQLPVILVGVKRCJINHAWL
 TQALIQHAGLTLACVAVANDVTPPEKRAEYITTLRMIAPRLGELIPMAENPE"
 BASE COUNT 1274 a 1507 c 1178 t
 ORIGIN
 Query Match 96.0%; Score 24; DB 2; Length 5526;
 Best local similarity 100.0%; Pred. No. 0.00079;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catgcacagcatlccgacccgct 25
 Db 2053 CATGCCAGCATTCGATTCGCT 2076
 RESULT 16
 PCOBIOB
 LOCUS ECOBIOB 128 bp DNA BCT 03-JAN-1995
 DEFINITION Escherichia coli biotin (bioB) gene, early terminator region.
 ACCESSION M27731
 VERSION M27731.1 GI:341755
 KEYWORDS bioB gene; biotin.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 128)
 AUTHORS Nath, S. K
 TITLE Attenuation of transcription of biotin genes in Escherichia coli
 JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
 MEDLINE 89167942
 FEATURES
 source
 1..128
 Location/Qualifiers
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 4..128
 /gene="bioB"
 4..9
 /gene="bioB"
 20..>128
 /gene="bioB"
 103..121
 terminator
 -10_signal
 mRNA
 terminator

BASE COUNT 31 a 28 c 34 g 35 t
 ORIGIN /gene="biog"
 /note="early terminator"

Query Match 72.0%; Score 18; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 cagattcgcattcgt 25
 |||||||
 Db 57 CAGATTTCGATCTCTCT 74

RESULT 17
 LOCUS G01446 479 bp DNA STS 19-AUG-1999
 DEFINITION Dm0466 Drosophila P1 library Drosophila melanogaster STS genomic
 clone DS07967 SP6, sequence tagged site.
 ACCESSION G01446
 VERSION G01446.1 GI:684849
 KEYWORDS STS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 479)
 RUBIN, G.
 TITLE Berkeley Drosophila Genome Project
 JOURNAL Unpublished (1994)
 COMMENT

Contact:
 Berkeley Drosophila Genome Project
 Primer A: CTCCTACCTGGGACATTCGTA
 Primer B: CACCAAGGCGCATATCCGA
 STS size: 192
 PCR Profile:
 Annealing: 58 degrees C PCR cycles: 30
 Protocol:
 Template: P1 Library Pools
 Primer: 1 uM each
 dNTPs: 250 uM each
 Taq Poly: 0.05 units/uL
 Total Vol: 15 uL

Buffer: MgCl2: 1.5mM
 KCl: 50 mM
 Tris-HCl: 50 mM
 pH: 8.3
 Gelatin: .001 %

The P1 library has been distributed to 16 regional sites. A list
 of these sites is available from FlyBase, via anonymous ftp to
 ftp.bio.indiana.edu in the file
 flybase/allied-data/genome-projects/1bl/LBL.doc.

FEATURES

source
 1. 479
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="DS07967"
 /note="Vector: Pad10sacII: The P1 library was made by D.
 Smoller in D. Hartl's lab (see Smoller et al., Chromosome
 100: 487). Chromosomal position in the Hartl lab (see
 Hartl et al., PNAS 91: 6824). STS's were generated by
 sequencing the ends of the Drosophila insert in these P1
 clones, by the W. Kimmerly, C. Martin, and M. Palazzolo
 lab at LBL."
 <1..250

primer_bind complement(233..250)
 BASE COUNT 113 a 120 c 97 g 145 t 4 others
 ORIGIN

Query Match 68.0%; Score 17; DB 54; Length 479;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gcatcgccagcattcg 17
 |||||||
 Db 396 GCATCGCCAGCATTTTCG 412

RESULT 18
 LOCUS AC019940 53067 bp DNA HTG 03-JAN-2000
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
 pieces.
 ACCESSION AC019940
 VERSION AC019940.1 GI:6664957
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 53067)
 ADAMS, M. and VENTER, J. C.
 TITLE Direct Submission
 JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 This sequence was identified as CDM:10211463 by the submitter.
 For more information on this record e-mail to flycelera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

1..53067
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 14944 a 11879 c 11720 g 14524 t
 ORIGIN

Query Match 68.0%; Score 17; DB 65; Length 53067;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcg 17
 |||||||
 Db 13629 GCATCGCCAGCATTTTCG 13613

RESULT 19
 LOCUS AC013431 153870 bp DNA HTG 31-JAN-2000
 DEFINITION Drosophila melanogaster chromosome X clone BACR22H11 (D1191)
 RPCL-98 22 H.11 map 13E-13F strain y: cn bw sp, *** SEQUENCING IN
 PROGRESS ***. 73 unstranded pieces.

ACCESSION AC013431
 VERSION AC013431.8 GI:6838815
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 153870)
 CEINIKER, S. E., AGAYANI, A., ARCAINA, T. T., BAXTER, E., BLASEJ, R. G.,
 BUTENHOFF, C., CHAMPE, M., CHAVEZ, C., CHEW, M., CIESLOJKA, L.,
 DOYLE, C. M., FARLAN, D. E., GALLE, R., GEORGE, R. A., HARRIS, N. L.,

TITLE
JOURNAL
AUTHORS
REFERENCE

Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshirefi, A.R.,
Moshirefi, M., Nixon, K., Paclet, J.M., Park, S., Pfeiffer, B.,
Richards, S., Sethi, H., Svirska, R.R., Man, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of *Drosophila melanogaster*
Unpublished
2 (bases 1 to 153870)

TITLE
JOURNAL
COMMENT

Submitted (11-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jan 31, 2000 this sequence version replaced q1:6532028.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to hdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
996 1075: contig of 995 bp in length
1076 1560: contig of 485 bp in length
1561 1640: gap of unknown length
1641 2230: contig of 590 bp in length
2231 2310: gap of unknown length
2311 2956: contig of 646 bp in length
2957 3036: gap of unknown length
3037 3621: contig of 585 bp in length
3622 3701: gap of unknown length
3702 4535: contig of 834 bp in length
4536 4615: gap of unknown length
4616 5559: contig of 944 bp in length
5560 5639: gap of unknown length
5640 6535: gap of unknown length
6536 6615: gap of unknown length
6616 7660: contig of 1045 bp in length
7661 7740: gap of unknown length
7741 8730: contig of 990 bp in length
8731 8810: gap of unknown length
8811 9404: contig of 594 bp in length
9405 9484: gap of unknown length
9485 10725: contig of 1241 bp in length
10726 10805: gap of unknown length
10806 11578: contig of 773 bp in length
11579 11658: gap of unknown length
11659 12834: contig of 1176 bp in length
12835 12914: gap of unknown length
12915 13501: contig of 587 bp in length
13502 13581: gap of unknown length
13582 14462: contig of 881 bp in length
14463 14542: gap of unknown length
14543 15566: contig of 1024 bp in length
15567 15646: gap of unknown length
15647 16366: contig of 720 bp in length
16367 16446: gap of unknown length
16447 17301: contig of 855 bp in length
17302 17381: gap of unknown length
17382 18390: contig of 1009 bp in length
18391 18470: gap of unknown length

18471 19409: contig of 939 bp in length
19410 19489: gap of unknown length
19490 20726: contig of 1237 bp in length
20727 20806: gap of unknown length
20807 22052: contig of 1246 bp in length
22053 22132: gap of unknown length
22133 23015: contig of 883 bp in length
23016 23096: gap of unknown length
23097 24622: contig of 1527 bp in length
24623 24702: gap of unknown length
24703 26689: contig of 1987 bp in length
26690 26769: gap of unknown length
26770 27825: contig of 1056 bp in length
27826 27905: gap of unknown length
27906 29904: contig of 1999 bp in length
29905 29984: gap of unknown length
29985 31588: contig of 1604 bp in length
31589 31668: gap of unknown length
31669 33112: contig of 1444 bp in length
33113 33192: gap of unknown length
33193 35374: contig of 2182 bp in length
35375 35454: gap of unknown length
35455 37223: contig of 1769 bp in length
37224 37303: gap of unknown length
37304 38978: contig of 1675 bp in length
38979 39058: gap of unknown length
39059 41337: contig of 2279 bp in length
41338 41417: gap of unknown length
41418 42461: contig of 1044 bp in length
42462 42541: gap of unknown length
42542 45823: contig of 3282 bp in length
45824 45903: gap of unknown length
45904 47247: contig of 1344 bp in length
47248 47327: gap of unknown length
47328 50153: contig of 2826 bp in length
50154 50233: gap of unknown length
50234 52182: contig of 1949 bp in length
52183 52262: gap of unknown length
52263 56756: contig of 4494 bp in length
56757 56836: gap of unknown length
56837 63525: contig of 6689 bp in length
63526 63605: gap of unknown length
63606 71374: contig of 7669 bp in length
71375 71354: gap of unknown length
71355 78141: contig of 6787 bp in length
78142 78221: gap of unknown length
78222 86570: contig of 8349 bp in length
86571 86650: gap of unknown length
86651 96227: contig of 9577 bp in length
96228 96307: gap of unknown length
96308 107431: contig of 11124 bp in length
107432 107511: gap of unknown length
107512 121618: contig of 14107 bp in length
121619 121698: gap of unknown length
121699 139797: contig of 18099 bp in length
139798 139877: gap of unknown length
139878 140388: contig of 511 bp in length
140389 140468: gap of unknown length
140469 141011: contig of 543 bp in length
141012 141091: gap of unknown length
141092 141513: contig of 422 bp in length
141514 141593: gap of unknown length
141594 142057: contig of 464 bp in length
142058 142137: gap of unknown length
142138 142690: contig of 553 bp in length
142691 142770: gap of unknown length
142771 143944: contig of 574 bp in length
143945 143424: gap of unknown length
143425 143857: contig of 433 bp in length
143858 143937: gap of unknown length
143938 144603: contig of 666 bp in length
144604 144683: gap of unknown length
144684 145116: contig of 433 bp in length

145117 145196: gap of unknown length
 * 145197 145734: contig of 538 bp in length
 * 145735 145814: gap of unknown length
 * 145815 146224: contig of 410 bp in length
 * 146225 146304: gap of unknown length
 * 146305 146859: contig of 555 bp in length
 * 146860 146939: gap of unknown length
 * 146940 147585: contig of 646 bp in length
 * 147586 147665: gap of unknown length
 * 147666 148081: contig of 416 bp in length
 * 148082 148161: gap of unknown length
 * 148162 148539: contig of 378 bp in length
 * 148540 148619: gap of unknown length
 * 148620 148950: contig of 331 bp in length
 * 148951 149030: gap of unknown length
 * 149031 149375: contig of 345 bp in length
 * 149376 149455: gap of unknown length
 * 149456 149853: contig of 398 bp in length
 * 149854 149933: gap of unknown length
 * 150934 150532: contig of 599 bp in length
 * 150533 150612: gap of unknown length
 * 150613 151147: contig of 535 bp in length
 * 151148 151227: gap of unknown length
 * 151228 151577: contig of 350 bp in length
 * 151578 151657: gap of unknown length
 * 151658 151964: contig of 307 bp in length
 * 151965 152044: gap of unknown length
 * 152045 152629: contig of 585 bp in length
 * 152630 152709: gap of unknown length
 * 152710 153245: contig of 536 bp in length
 * 153246 153325: gap of unknown length
 * 153326 153870: contig of 545 bp in length.
 Location/Qualifiers
 1. 153870
 /organism="Drosophila melanogaster"
 /strain="Y: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="X"

FEATURES
 source
 Query Match 68.0%; Score 17; DB 63; Length 153870;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcacgacattcg 17
 |||||
 Db 63673 GCATCGCCAGCATTCG 63689

RESULT 20
 AC007809 183439 bp DNA INV 24-FEB-2001
 LOCUS Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone
 BACR45M03, complete sequence.
 AC007809
 AC007809.8 GI:13122709
 HTG.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 183439)
 Celinker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanalides,P.G., Brandon,R.C.,
 Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
 Carlson,J.W., Center,A., Chapple,M., Davenport,L.B., Dietz,S.M.,
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
 Ferreira,S., Frisoe,E., Galle,R.F., Gary,N.S., George,R.A.,
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
 Ileguam,C., Jalili,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
 Paaleh,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,

Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
 Sequencing of Drosophila chromosome 3R, region 88C-88C
 Unpublished
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

2 (bases 1 to 183439)
 Celinker,S.E., Agbayani,A., Arcaluna,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Chapple,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummel,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Paaleh,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snit,E.,
 Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 24, 2001 this sequence version replaced g1:6623906.
 COMMENT
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.

FEATURES
 source
 1. 183439
 /organism="Drosophila melanogaster"
 /strain="Y: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="88C-88C"
 /clone="BACR45M03 (D718)"
 /clone_id="RPC1-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 PBAC3.6)"

BASE COUNT 53300 a 38849 c 38854 g 52436 t
 ORIGIN

Query Match 68.0%; Score 17; DB 4; Length 183439;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcacgacattcgatc 20
 |||||
 Db 66707 TCGCCAGCATTTGATC 66723

RESULT 21
 AE003705 232744 bp DNA INV 05-OCT-2000
 LOCUS Drosophila melanogaster genomic scaffold 142000013386035 section 30
 DEFINITION of 105, complete sequence.
 ACCESSION AE003705 AE002708
 VERSION AE003705.1 GI:7299886
 KEYWORDS
 SOURCE
 ORGANISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 232744)
 Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
 Britton,R.C., Rogers,J.H., Blazej,R.G., Chapple,M., Pfeiffer,B.D.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
 Miklos,G.L., Adril,J.F., Agbayani,A., An,H.J.,

Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
 Baxendale,J., Bayraktaroglu,I., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borikova,D.,
 Bouchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtils,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrer,S.,
 Gelbach,W.M., Glasser,K., Glodex,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpn,G.H., Ke,Z.,
 Krenitsen,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Mekulov,G., Mishina,N.V., Mobarry,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,D., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
 Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 232744)
 Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

FEATURES
 source
 Location/Qualifiers
 1..232744
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3R"
 complement(join(29274..29439,29524..30038,30107..>30980))
 /gene="CG8464"
 /product="CT24781"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 complement(<29274..>30980)
 /gene="CG8464"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 /note="CG8464 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 /protein_id="AAF5062.1"
 /db_xref="GI:7299887"
 /translation="MALRGSRLLEVIFKRCIASPVLSHAANRSSOLAIREGDPNSN
 GNSGOYQNGEOKERKRLVRFVPSGLGAVSAIIQREDDLPITIAASKMGGRRD
 FNFIADVAGGACDSVYIEIKDRHDFVSGPITANSNGSITFQNLILITNAIVY
 NKPHIVGVRLSDGRFPATIEDVDOTSLATLRQVNNLSMRIGKSTLRSGEYV
 AIGSPILASNTYTAGVISTQASOELGRNNDIYDITDAITFGNSGGPILVNDGE
 AIGVSMKVTAGISFAIPIDYKVLERAERKRKGSAYKGYPKRYMGITMLITLP
 DILFELSKSQMPPSNLTHGVLMKVIAGVSPAHSGQLQPCDITVTHINKKEIKNSDY
 DALADNSKTLIDVILRGVGMHVTITPDP"
 join(<31212..31346,31407..31691,31752..>31922)
 /gene="CG3351"
 /product="CT11255"

gene
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 <31212..>31922
 /gene="CG3351"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 join(31212..31346,31407..31691,31752..31922)
 /gene="CG3351"
 /note="CG3351 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 /protein_id="AAF5063.1"
 /db_xref="GI:7299888"
 /translation="MSKAAGKLSUKTVERVYTSKLTNIIPAGMAAGPPLGMLG
 ORAINIAFCQDNFAKTAEMKGVPLPQRIVSNSDRSYDLAIIHPDPATFLKQAGIG
 RQTMTPGKEVAGMITLKILYETALIKIDPPNVLITQMOCEMLISTARTGIVYRE
 IDPAVGEFLERKTLIVEQORELEQERKARMLRPG"
 complement(join(32136..32708,32770..>32889))
 /gene="CG8461"
 /product="CT24741"
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 complement(<32136..>32889)
 /gene="CG8461"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 complement(join(32136..32708,32770..32835))
 /gene="CG8461"
 /note="CG8461 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 /protein_id="AAF5064.1"
 /db_xref="GI:7299889"
 /translation="MSSPSSSSASAEHSDNETONAIREDLKGSFEEIMLKLEEL
 GAKYKKEAVLGSKSSRPQPKAKTDLRLKLNRPREMTTRQVPEFLAEHRVERKKVL
 ELRDRPREDEKSGTSAETPKKQYQVFOVSRKEVGLKKLDEVEDAEKHHIKDTMV
 RLINKNVDEKKHKKOKOLKREKRSALIKKHDLGQDPHYLRKSPFLIDSEIKF"
 complement(join(<34919..35540,35598..35757,35894..36047,
 36334..36679,36886..37166,37233..37324,37381..37624,
 37694..37981,38058..38138,42442..42707,42765..43010,
 44497..44761,48465..>49118))
 /gene="CG8279"
 /product="CT24224"
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 complement(<34919..>49118)
 /gene="CG8279"
 /map="88C5-88C6"
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 complement(join(34919..35540,35598..35757,35894..36047,
 36334..36679,36886..37166,37233..37324,37381..37624,
 37694..37981,38058..38138,42442..42707,42765..43010,
 44497..44761,48465..49118))
 /gene="CG8279"
 /note="CG8279 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 /protein_id="AAF5066.2"
 /db_xref="GI:10726529"
 /translation="MHGPVSRSSSSNMVDVSSPAGGAAPVEMSTTSSSAATTSAS
 SSKPLTNGAKTIAISTAGGVTPGAVGPCSGAIPASSSSGNQVLEHNRVENNR
 AVINRSSETILMTPTGSSSPQSPGOTASLOTQNSODRLAKASTESODVDYVA
 RLPEERKEAPEKLTERRAPREALSRQERLENNKPKRVSYSDLFOQMAASPTVOO
 KSRFSLSNASSSLPECRHMDLDEGELEPMELIRVDVANEIDVLCRILVNVGLT
 HADRSGLFLAKGTPTNKYVIAKLEVDYQALAKLDVATRSABEIIIPFAGIGAGVQO
 TKOMINIKAVYKDAFENCETDLTKYRTNAILCMPCNCEGDIIGVQILNTKNGME
 FDEHVEIERRYLTFPGIGIQNQLFEMSVQEVYRRNQDILNLARSIFEDQNNDECLVT

KIMTARELLKCCRSVFLVDLCCASHLEKTEKPNOPATRAIKSADSEKKMRN
 RETVFLGEGEAAANVSPPSVSELSSTLQIAOFVATGCTVACIVIEVDRHNO
 IRAEDELSTOAILCMPIINMAKRVIGVAGOLINKANGVPTDSASIEFAFIPGGLG
 IINTOMYENACKLMAKOKVALECLSYHATASODOTEKLTODVIAEASVNSXSPFED
 FELVDDTCTRAVLRRPMOCNIVSOPITPDVLCRVNVSRYKRYKRVKHNHNLNVA
 QTFMFLKTKGRMERMTDEILGLVACICHDHGRGNMOTTEPPLALVLTST
 MEHHHDEHRLAFLDLRAAFRRALRLARLDELIVELVNDPTLPIVIEFRLRR
 TLEFPELDEEDDDVDVSVVVLVSVLVDSVLVLSALTWATYKTIERTKRT
 RDNFMDALSPEDYSKMTVESALISTDLADFEKKNAPLELVENGSEFDMGEEKRD
 LLCGMATACDVSALAKPHEVOHKAALVADFEPOGDLERQILNTOVAMANDREKRD
 ELPKMGVFIIVICLPYLRVLCDFPWTPLTEGTVLENRMODLAERVEKLTWIDH
 DTIDRVEFEACADEIKDIEFTVTLNCSQSHGSEDSITPEHOSGSLSKMT
 GALGAVRSLKSLTYNSMDGSKPKTSKLESHVSEDMDKSPSPSQPOASGSMGR
 MSASSTSSAGGOMDKSKRSKLCALL"
 join(40359, 40547, 40626, 40826, 40891, 41076, 41138, 41210,
 41284, 41410, 41461, 42034, 42102, 42262)
 /gene="Cyp313a1"
 /product="Cp311271"
 /db_xref="FLYBASE:FBgn00038236"
 /db_xref="FLYBASE:FBgn00038236"
 <40359, >42262
 /gene="Cyp313a1"
 /note="CG3360"
 /map="88C5-88C6"
 /db_xref="FLYBASE:FBgn0003360"
 /db_xref="FLYBASE:FBgn00038236"
 join(40391, 40547, 40626, 40826, 40891, 41076, 41138, 41210,
 41284, 41410, 41461, 42034, 42102, 42262)
 /gene="Cyp313a1"
 /note="Cyp313a1 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBgn0003360"
 /db_xref="FLYBASE:FBgn00038236"
 /protein_id="AAF55065.2"
 /db_xref="GI:10726530"

Query Match 68.0%; Score 17; DB 5; Length 232744;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tgcgcagcattcgatc 20
 |||||
 Db 106438 TCGCCAGCATTCGATC 106454

RESULT 22
 AC017740/c DNA HTG 10-DEC-1999
 LOCUS
 DEFINITION AC017740 237119 bp
 AC017740 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 VERSION AC017740.1 GI:6554259
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 237119)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10211941 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 SOURCE
 1. 237119
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 67689 a 49909 c 50172 g 69349 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 65; Length 237119;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tgcgcagcattcgatc 20
 |||||
 Db 103681 TCGCCAGCATTCGATC 103665

RESULT 23
 AE003500
 LOCUS
 DEFINITION AE003500 Drosophila melanogaster genomic scaffold 14200001336053 section 17
 ACCESSION AE003500 AE002593
 VERSION AE003500.2 GI:10728273
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 327446)
 Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
 Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Gall, R.F.,
 George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
 Sutton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
 Burton, R.C., Rogers, Y.H., Blazer, R.G., Chapple, P., Pfeiffer, B.D.,
 Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gaber
 Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
 Andrews, P., Binkov, C., Baldwin, D., Balen, R.M., Basu, A.,
 Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
 Benos, P.V., Bernier, B.P., Bhattacharya, D., Bolshakov, S., Borokov, B.C.,
 Boudreau, M.R., Bouck, J., Brokstein, P., Brotlier, P., Burris, K.C.,
 Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
 Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
 Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
 Fleischmann, W., Folsler, C., Gabrielian, A.E., Garg, N.S.,
 Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorelli, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, D.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Pacle, J.M., Palazolo, M., Pittman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Rehrer, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shu, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svatek, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
 Weinstein, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 Zaveri, J.S., Zhan, M., Zhang, S., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of Drosophila melanogaster
 Science 287 (5461), 2185-2195 (2000)
 2 (bases 1 to 327446)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

oy	1	gcatacgccagcatctg 17
Db	70808	GCATCGCCAGCATTTCCG 70824
RESULT 24		
AF146614/c		
LOCUS		
DEFINITION	AF146614 1186 bp DNA BCT 09-AUG-1999	
	Erwinia carotovora subsp. carotovora xerD (xerD) gene, partial cds;	
	Dsbc precursor (dsbc) gene, complete cds; and RecJ (recJ) gene,	
	partial cds.	
ACCESSION	AF146614	
VERSION	AF146614.1 GI:5712695	
KEYWORDS		
SOURCE		
ORGANISM	Pectobacterium carotovorum subsp. carotovorum, Pectobacterium carotovorum subsp. carotovorum, Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.	
REFERENCE	1 (bases 1 to 1186)	
AUTHORS	Vincent-Sealy,L.V., Thomas,J.D., Commander,P. and Salmond,C.P.	
TITLE	Erwinia carotovora dsbc mutants: evidence for a periplasmic-stress signal transduction system affecting transcription of genes encoding secreted proteins Microbiology 145 (Pt 8), 1945-1958 (1999)	
JOURNAL	99392457	
MEDLINE	2 (bases 1 to 1186)	
REFERENCE	Thomas,J.D.	
AUTHORS	Direct Submission	
TITLE	Submitted (23-Apr-1999) Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK	
JOURNAL	Location/Qualifiers	
FEATURES		
source	1. .1186	
	/organism="Pectobacterium carotovorum subsp. carotovorum"	
	/strain="SCRI193"	
	/sub_species="Carotovora"	
	/db_xref="taxon:355"	
	<1. .189	
gene	/gene="xerD"	
	<1. .189	
CDS	/gene="xerD"	
	/note="similar to Escherichia coli XerD integrase/recombinase in SwissProt Accession Number P21891"	
	/codon_start=1	
	/transl_table=11	
	/evidence=not experimental	
	/product="xerD"	
	/protein_id="AAD47611.1"	
	/db_xref="GI:5712697"	
	/translation="KLSPIYLRHAFATLLNMGAGFTGRTDALGQRSFHDADLPCH DGAETRASAAHVRHRI"	
gene	222. .1031	
	/gene="dsbc"	
	222. .227	
-35_signal	/gene="dsbc"	
	246. .251	
-10_signal	/gene="dsbc"	
	258. .264	
RBS	/gene="dsbc"	
	/note="putative"	
	315. .377	
sig_peptide	/gene="dsbc"	
	/note="putative sec-dependent export signal"	
	315. .1031	
CDS	/gene="dsbc"	
	/function="periplasmic disulfide bond formation"	
	/note="disulfide isomerase; similar to Erwinia chrysanthemi and Escherichia coli Dsbc proteins in SwissProt Accession Numbers P39691 and P21892"	
	/codon_start=1	
	/transl_table=11	
	/product="Dsbc precursor"	

```

/protein_id="AAD47610.1"
/db_xref="GI:5712696"
/translation="MKKGILLISLVATATNFNPAHADDAIKOTLHAGVARGGGPAPAA
DGMKRIIVTDSGVLIYISEDGQAFRAPALRRERTPVNTTKILIGLDLKLODMOVI
KAQEKIIVFTSITCGCYCHKLHEOKMDNALGITVRILAEPQGCKSPAAKDMSY
WCVADEKRKAFFDSAMRGDAISAATCTDIAAHYOLGFQVGQPPIAVLQGLVPYGVO
GPKEMLAMLFAHKASKKTGS"
    mat_peptide
        378..1028
            /gene="dsbc"
            /product="Dsbc"
            /gene="dsbc"
            669...680
            /gene="recJ"
            /note="Region: putative CPHC motif typical of the active sites of disulfide isomerases"
            1044..>1186
            /gene="recJ"
            1044..>1186
            /gene="recJ"
            /note="theoretical protein; similar to Erwinia chrysanthemi and Escherichia coli RecJ single-stranded DNA-specific exonucleases in Swissprot Accession Numbers P39693 and P21893"
            /codon_start=1
            /transl_table=1
            /evidence=not experimental
            /product="RecJ"
            /protein_id="AAD47612.1"
            /db_xref="GI:5712698"
            /translation="MITQLRRRPLAEIDIDPTFVPPLLRLYLHAHGKYGAOLERSLS
            GLL"
BASE COUNT      295 a       315 c       323 g       253 t
ORIGIN
Query Match          64.0%; Score 16; DB 2; Length 1186;
Best Local Similarity 100.0%; Pred. No. 41;
Matches   16; Conservative % 0; Mismatches     0; Indels    0; Gaps    0;

OY      1 gcattgcacgaatctc 16
         |||||||iiiiiiiii
Db       993 GCATGCACGCAATTTC 978

RESULT_25
AF206717/c      SYN      12-SEP-2000
LOCUS           AF206717.1 GI:7262998
DEFINITION      Shuttle vector p13 hypothetical proteins, Htru (htru), Ardu (ardu)
                RepU (repU), and Resu (resu) genes, complete cds.
ACCESSION       AF206717.1
VERSION         AF206717.1
KEYWORDS        .
SOURCE          Shuttle vector p13.
ORGANISM        Shuttle vector p13
                artificial sequence; vectors.
REFERENCE       1 (bases 1 to 11910)
AUTHORS        Meima,R. and Lidstrom,M.E.
TITLE           Characterization of the minimal replicon of a cryptic Deinococcus
                radiodurans SARk plasmid and development of versatile Escherichia
                coli-D. radiodurans shuttle vectors
FEATURES             1..11910
                     JOURNAL APPL. Environ. Microbiol. 66 (9), 3856-3867 (2000)
                     MEDLINE 20422197
                     PUBMED 10966401
                     REFERENCE 2 (bases 1 to 11910)
                     TITLE Direct Submission
                     AUTHORS Submitted (18-NOV-1999) Chemical Engineering, University of Washington, Benson Hall, PO Box 351750, Seattle, WA 98195, USA
FEATURES
source          Location/Qualifiers
                 1..11910
                   /organism="Shuttle vector p13"
                   /specific_host="Deinococcus radiodurans SARk"
                   /db_xref="taxon:115340"
                   /note="derived from pUE10"
```



```

repeat_region /note="L1 repeat: matches 4449. .5390 of consensus"
40148. .41111
/note="L1MD1 repeat: matches 1. .968 of consensus"
repeat_region complement(41972. .42157)
/note="MLTIF repeat: matches 541. .354 of consensus"
repeat_region complement(43606. .43671)
/note="MIR2 repeat: matches 134. .71 of consensus"
45479. .43585
/note="MER45 repeat: matches 26. .133 of consensus"
48242. .48312
/note="MER45 repeat: matches 98. .176 of consensus"
49604. .49826
/note="AluB repeat: matches 82. .301 of consensus"
49936. .50472
/note="L1PA13 repeat: matches 41. .573 of consensus"
complement(50932. .51134)
/note="MER20 repeat: matches 218. .18 of consensus"
51378. .51596
/note="MER20 repeat: matches 1. .215 of consensus"
complement(52293. .52552)
/note="MLTIC repeat: matches 466. .186 of consensus"
repeat_region complement(52555. .52908)
/note="THBIB repeat: matches 362. .1 of consensus"
complement(52914. .53071)
/note="MLTIC repeat: matches 182. .8 of consensus"
53212. .53621
/note="WSTC repeat: matches 1. .403 of consensus"
55202. .55295
/note="MIR2 repeat: matches 1. .94 of consensus"
55296. .55504
/note="MER3 repeat: matches 1. .209 of consensus"
complement(55783. .56684)
/note="L1PA10 repeat: matches 911. .1 of consensus"
complement(56534. .57144)
/note="L1 repeat: matches 5390. .4785 of consensus"
complement(57148. .59003)
/note="L1 repeat: matches 4720. .2850 of consensus"
59735. .59757
61714. .61945
/note="match: STS L24834"
complement(63228. .63358)
/note="MIR repeat: matches 247. .100 of consensus"
repeat_region complement(64335. .64455)
/note="L1MD3 repeat: matches 909. .790 of consensus"
complement(65455. .65744)
/note="AluB repeat: matches 289. .2 of consensus"
66343. .66557
/note="MIR repeat: matches 41. .261 of consensus"
66912. .67656
/note="L1MD3 repeat: matches 1660. .2409 of consensus"
complement(68490. .68794)
/note="AluB repeat: matches 301. .1 of consensus"
complement(68873. .69190)
/note="AluB repeat: matches 298. .1 of consensus"
69217. .69332
/note="MER33 repeat: matches 1. .324 of consensus"
complement(69638. .69703)
/note="MIR repeat: matches 129. .64 of consensus"
71210. .71340
/note="MIR repeat: matches 83. .206 of consensus"
72426. .72792
/note="match: GSS A0053969"
complement(72923. .72973)
/note="MIR repeat: matches 139. .89 of consensus"
complement(75007. .75196)
/note="MER5A repeat: matches 189. .1 of consensus"
complement(75386. .75425)
/note="L1MD3 repeat: matches 406. .367 of consensus"
75508. .75800
/note="AluB repeat: matches 1. .288 of consensus"
complement(76744. .76840)
/note="MLTIF repeat: matches 182. .91 of consensus"
complement(78342. .78521)

```

```

repeat_region /note="MIR repeat: matches 216. .35 of consensus"
complement(78988. .79020)
/note="MIR2 repeat: matches 144. .112 of consensus"
repeat_region complement(79137. .79186)
/note="MIR2 repeat: matches 134. .85 of consensus"
repeat_region complement(79151. .79393)
/note="MIR repeat: matches 242. .3 of consensus"
complement(79492. .79666)
/note="MER5B repeat: matches 178. .5 of consensus"
complement(79777. .79863)
/note="MER5A repeat: matches 109. .27 of consensus"
79894. .79970
/note="MER5A repeat: matches 11. .91 of consensus"
complement(80377. .80431)
/note="MIR2 repeat: matches 141. .87 of consensus"
complement(80389. .80462)
/note="MIR repeat: matches 252. .177 of consensus"
complement(80455. .80578)
/note="MIR repeat: matches 147. .29 of consensus"

```

```

Query Match 64.0%; Score 16; DB 92; Length 117968;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 atgcgcagattccga 18
|||||
Db 97872 ATGCCAGATTCCGA 97857

```

```

RESULT 27
AC017911 118100 bp DNA HTG 09-DEC-1999
LOCUS AC017911
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017911
VERSION AC017911.1 GI:6553279
KEYWORDS HTG; HTGS; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 118100)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Substitution
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

```

```

COMMENT This sequence was identified as CDM:10212606 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

```

FEATURES
source 1. 118100
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

```

```

BASE COUNT 32824 a 26059 c 26550 g 32667 t
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 65; Length 118100;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 gcaattgcattccgct 25
|||||
Db 63425 GCATTTCGATCCGCT 63440

```

```

RESULT 28
AP002817/c
LOCUS AP002817 140825 bp DNA PLN 12-AUG-2000

```

DEFINITION	Oryza sativa genomic DNA, chromosome 1, PAC clone:P0699D11.	CDS	GSDDCLADAASYLGAELORURPRCG"
ACCESSION	AP002817 BA000010		complement(join(15587..15727,15868..15915))
VERSION	AP002817.1 GI:9558510		/note="hypothetical protein"
KEYWORDS			/codon_start=1
SOURCE	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0699D11.		/protein_id="BAB03431.1"
ORGANISM	Oryza sativa		/db_xref="GI:9558513"
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.		/translation="MNNFHVVEAIGRGKHSTVYGRKKKSIIEYFAVKSVDKQSRSKVL
REFERENCE	1 (bases 1 to 140825)		NEVSAAPSSPCSFISSSW"
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		complement(join(18733..18801,19343..19489,19843..19905,
TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0699D11		20016..20087,20611..20691,20769..20819,20976..21072,
JOURNAL	Published Only in DataBase (2000) In press		21611..21753))
REFERENCE	2 (bases 1 to 140825)		/note="ESTS C72327(E1437),A0078752(E1437) correspond to a
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		region of the predicted gene.
TITLE	Direct Submission		Similar to Archaeoglobus fulgidus competence-damage
JOURNAL	Submitted (26-JUL-2000) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan		protein (AE000949.)
REFERENCE	(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)		/codon_start=1
COMMENT	The orientation of the sequence is from sp6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP(PIR,SWISSPROT, GENPEPT, PDB) from MAFF DNABank and the cDNA sequence database at RGP. Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DDBJ accession no. and RGP clone ID.		/protein_id="BAB03432.1"
	This sequence of P0699D11 clone has an overlap with P0462H08 clone, DDBJ:AP002525 at the 5' end. The sequence of this clone starts at the position 114,739 of P0462H08. This sequence of P0699D11 clone has an overlap with P0469E09 clone, DDBJ:AP001366 at the 3' end. The sequence of this clone ends at the position 59,396 of P0469E09. Detailed information on assemble quality together with annotation of this entry at		/db_xref="GI:9558514"
FEATURES	http://rgp.dna.affrc.go.jp/GenomeSeq.html.		/translation="MISDGRLEAGRTKKNISVSSNNTNSTEVEQTISRASIIIVG
SOURCE	Location/Qualifiers		DEITFTQVEDKLGAGLCKLHAIGWRVSHVAVVNEIDSAEVEVERCKSTDMMVFLVG
	1. 140825		GLGLHSDISLAGAKAFQVRLAPDEFEYLSQIDGNTGDRNEMALLPGITELL
	/organism="Oryza sativa"		HKMLPPLIKCKNVVLLAATNVDELETEMGCLDTQESGLVMAKSFVSKHLCITSLID
	/cultivar="Nipponbare"		VKIAPVVKLCIDFSDVYIGMI"
	/db_xref="taxon:4530"		join(30551..30609,30919..31139,31223..31578)
	/chromosome="1"		/note="Similar to Arabidopsis thaliana chromosome 4, BAC
	/clone="P0699D11"		clone F10W23; hypothetical protein (AL035440)"
	join(527..679,925..1114,2076..2467,3067..3555)		/codon_start=1
	/note="EST C22619(S11214) corresponds to a region of the predicted gene.		/protein_id="BAB03433.1"
	Similar to Arabidopsis thaliana chromosome 2, BAC clone T17D12; putative protein kinase (AC006587)"		/db_xref="GI:9558515"
	/codon_start=1		/translation="MUCGKRGRRREGGNEDDIVAKKRSMGCPDPLHLRCKSISKVP
	/protein_id="BAB03429.1"		OLVIGIKNVLEPNRYAGSLNCTKPVTRGSSVVLTPVLVYTAESWPNRIVKRGHGH
	/db_xref="GI:9558511"		HGDIROQGAGAAERDEGVVTDGLLHQHQRPHVIPSEAAATVRRRGWATOSSVGED
	/translation="MVGIRNKGHEATIFTRELADATNMFSTECLLRGGRGVSKY		GDAAMVSPHKVVARRAAASHSSVLEGAGMLKGLDHLHVRNAVLCRTGFLD"
	AFLNDQVAVKQLDNLGLOGREFLVEVLMLSLLHHPNLVKLGYCVGDGDORLLIYE		complement(join(31912..32455,32698..33684,34091..34341))
	YMPGLSLEDRULHURPOGEPLDWTTRMKIAADAAAGLEYLHDEAIFAVYIRKIPNSI		/codon_start=1
	LLGEGYNKLSDFGLAKLPGVDKTHVTVMGTHTGYCAPEYLSLTKLTKISDIYSFG		/protein_id="BAB03434.1"
	VVFLELITGRALDSNRPDEQDLVAMARPLFKDORRFPKADPSLHGHPKRGFLQA		/db_xref="GI:9558516"
	LAIAMCLOEKAKNRPSIREVAVALSYSATQESQNTAARHTLPGPSVPRVLDNQIN		/translation="MRKKRTNLLIIVGGEATFFFFEERCRRRAGVGSEVERRGDSKP
	QDTSLPQSHGVHMPPLAGTDHMQVEVENCSSSRHPRGRVTPNGADRERALEANY		SRGRHPPTARAAPLRREOKRAREDELATASSGTGPGSWPHHPLSPSSPILLSLH
	WVEAWRQEKTSKMR"		SLLFSRGLGHEGCGFAASAGESGASAGSAAAAAAGGCGGCGWGAAPTSP
	complement(8769..9326)		RHFALSSAAPGARDGADAGLVGGVGGGTGEAGARARRRQIGOERCHMGKAPSV
	/note="ESTS AU093260(C52190),A0062991(C52190) correspond to a region of the predicted gene.		VCCCFPTVAELVVLAVCVPAALCHLAVRGHRRVCSAKOKEMGELLTLDVTPRS
	hypothetical protein"		LAATAAKARKEVEFTPATMAELHGEAEVARKMSGLQALHWWVRLSSAGEDARPA
	/codon_start=1		DLRVLLPLACPLSTIAPDLPRPHVRSLSCLSPKRLASITLSIPFORKNGDPSRSH
	/protein_id="BAB03430.1"		TEGNIKERRENSITPAQKTCRSRSAPATPPTVAPRPLTNHPIRQAPQRCR
	/db_xref="GI:9558512"		EEGGGGGDAALPACLLAASLPKVPTEPTKPAVRPLTNHPIRQAPQRCR
	/translation="MDDFLEPTLELCRAFTIRIGTSSKIVALSKEPALLVDSAFMSI		SSPSLSSLSRSGSGKPARCCTVARERLFWRRVRLRATSIHAEARAAHVIVVLV
	AVDQSCVMDICDFGNGMGIFDLVGSDDPHISDLASDCLVLLKKAAPREATVGLLTN		VKLPPAREACPRSTAPASRQGPPEHRR"
	LPKLSVLLDLKHGTGTLRLTLRYCLAFSCRQYLAQGMIVSISLSALMRVEALVSFAK		join(35382..35472,36089..36172,36426..36530,37078..37337,
			38995..39114)
			/note="hypothetical protein"
			/codon_start=1
			/protein_id="BAB03435.1"
			/db_xref="GI:9558517"
			/translation="MQLDASDFHGOELTLKLRDGLKSLSSLESSGKSYDLVSFAQOOP
			DATVFLPSGPEAKAVGIARRVSLVYPDPEELEKPLGLSLTPSSRKSAGSKSRKSR
			FTSGSKNRSGSQAQSLGQSAFAHKNOKRKDESLGSHNSVSGKSAGSQVRGDS
			GTTSEVPTQPEVSKSKKKLDRPKRRTPYDRHKAQELDRATVAVDSSLYSPKPY
			Q"
			Join(40731..40753,43157..43538,45516..45821,47347..47893,
			48657..48934,48983..50395)
			/note="Similar to Arabidopsis thaliana chromosome II BAC
			clone T27A16; hypothetical protein (AC005496)"
			/codon_start=1
			/protein_id="BAB03436.1"
			/db_xref="GI:9558518"
			/translation="MYSLRKIRPPPPALAAARDSPLRWPPPPGRLPSWIRWGRDR
			RRLLPYAMSRGPPPEGSGRGRENEEERSRRGGGEGGEGEEIKGMGRKEEREKE
			ETRLCAGRMRRDACTGSFADFLSLFNKSLVSLWLTAKDEIKQLAGNRNRDPDITGG
			HGAADKDLPIORSIVEIKLAGHISIHOLASCWHLHGPCSSCAGLRWRKRSASAL
			LSLSCSWNPQFCQKPNRSORPATTLTSSLLILSPFFSYLFYFQNHYSIFTSI
			FSLNLFGRFRSPPTISIVRPTAMRTVRSIVTIPYVDGVYEWKRLAPYTSIAISMSS

SLNPNMQTVRCIGSSALFSLGIGTPIHRAAKPIAPFGYCNFPFFFLKKYSKTEALAH
 AYTAHSSINRAHTPIRKIDILYKOARNLCTSPCKNRSRHINISAOCHTNNRPPAQ
 YRVDSLSEARQREAREARETARPHDRTSSTENRPPPPATAERWPPPPPPPPPP
 DLKHAAYLLPRGHHRAVASALLASILRLPDHAAALVRRVPSVSLRAARLRG
 RGGGGLAOLHSLVLRAGHADPHASASILVQYSCGSASARVPRDETAASADVVSW
 NVMDIGVYSGDLARARELDFVWPGNRVSVMTVIGAYAKOPPEATEVFERMVEG
 IEPDGVALLSVLSACDGLGVLDGEWVHRVFLRGCOEIPLMNAIIDIMVYKSGVKK
 ALBVFGEQKQISIVTWTWMIAGFALHGLEAVELFRMEKENVSPNDITFLAVLSVC
 SHVGLDGLRWTFKPMVSOYKIKPRVHYKCGMIDLLGRAGCLMEARGLIDMPPFKANA
 ATWGLLAAATVHGDTCEQALLHLIELEPHNSGNYILLNIVAEQERWDVARELRI
 SMRDGLRNVPCASSIDVDMVHEPISRDSGSPSLHKREVLCAINSKNSVGHIALLL
 PESLHDVEEG

CDS

complement(join(52441..53547,53887..54014,54128..54165,
 54272..54487,54709..54870,55838..56016,56150..56256,
 56375..56539,56658..56741,56821..57040))
 /note="Similar to Arabidopsis thaliana chromosome I BAC
 clone T5A14; kinesin (AC005223)"
 /codon_start=1
 /protein_id="BAB03437.1"
 /db_xref="GI:9558519"

/translation="MDVQPARTMRNLPTLSLMGFNKLTPSWTESVSHIIDGLSPT
 KPMVMVEKDEISDDTSEAKYQKIDELVSLNAQLQITLQRRALNNYLDLKG
 NTRVFCRIPFHEESYSRNLFTIDESNVFLKVAETKRYKQYKFKVDQFSTQDVF
 SEVERVKSALDGVNVCIFYAGOTSGTKTYMEGKPTNLGVPICLOTLENOASECN
 RELFTFMLEIYMGIRDLAPRSKNGIKNVPSLIKSDPDGIEIDELVAVTVNSF
 QEVRLYEMGTLSRSTASTMANSTSSRSCHLIRISLTINATERKATSKLWMDLGG
 SERLVKATGKRLKEGRANLSALGDVIDALQTKKPHVYRNSKLTQVLRDLSGC
 ESKTLMVHISPDGDLCEITCTHGFATVRVRSIRLESSEPEMKARKETLLIDLGKV
 NDLEHEDDIRKINKLESEHGTLPQPTIYSNFDSSLEELTKARKETLLIDLGKV
 REASRLPRMPTASQSHRIGLNRRPIINRLKPPVPPRRPPSSVVAESVMVPA
 APWQSECSSEKSMSTDSMNWTPSIRDGTEQSQDASEVEIKQVIFSEHEKSSHDQV
 YTDYPLASRDIOIKIEKGIVDIDNLHOOIVKSTPFRSKMVIDIPGYEAEIHSV
 SIPSPTTWTACIKEDSOVKDVMGLTLQSTDIYVEDIKSGKDNQFTAKELCTPPKEF
 SSNNEVKGHKEHPYHGRPRRSQLEENCTLEKPNMDSKSHRSHDDKKTGNVLSY
 TGE"

CDS

join(57603..57645,57742..57843,58195..58344,59425..59479,
 59537..59589,61304..61785)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB03438.1"
 /db_xref="GI:9558520"

/translation="MKQDKGNRCLFQGSFAPYIHOHVTIKGEECTDTPSPERDSK
 SEYAKMIDYIPWMAITFGNMRCLANENHSVAVIRKGSPTMCHAELESEATAIS
 LFLASLAPKFLSKRRSGDVHLMASIGVEAAAPLPPOPLPPGDRMPAAGG
 LCLSPASRLRSPPLPRQOPPMRDAPAAAGLRLSPNARRREIKRPLPSTSP
 AIQPCRWPPPLPADPATGELELYSSKRCRSCVERIPTVELLPMVGELELFSSTVKL
 LPSPSPSPAMSLAPSSRA"

CDS

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9; hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

Query Match 64.0%; Score 16; DB 13; Length 140825;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 catcgccagcatttcg 17
 |||||
 Db 89622 CATCGCCAGCATTTCG 89607

RESULT 29
 AP001366/c
 LOCUS AP001366 146081 bp DNA PLN 12-AUG-2000
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0469E09.
 ACCESSION AP001366 BA000010
 VERSION AP001366.1 GI:7228436
 KEYWORDS
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0469E09.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 Oryza.
 1 (bases 1 to 146081)
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:P0469E09
 Published Only in DataBase (2000) In press
 2 (bases 1 to 146081)
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 Direct Submission
 Submitted (09-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji
 Sasaki, National Institute of Agricultural Resources, Rice Genome
 Research Program; Kannonndai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@agr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 Genes were predicted from the integrated results of the
 following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
 SplicePredictor (October 1998 version). The genomic sequence was
 searched against the non-redundant database NRP (PIR, SWISSPROT,
 GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at
 RGP. Protein similarities of the coding regions were searched
 against NRP with BLASTP2.0. ESTs represent the identified cDNA
 sequences using BLASTN 2.0 with the corresponding DDBJ accession
 no. and RGP clone ID.
 Detailed information on overlap and assemble quality together with
 annotation of this entry at
 http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.

FEATURES
 source

Location/Qualifiers
 1..146081
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0469E09"
 join(5902..5980,6229..6272,8134..8295)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92397.1"
 /db_xref="GI:7228437"
 /translation="MCAVKRRSFLVVVVVFAILLTMAAGRKMLINKHQVQSMETS
 DESMHQGGDEDEMLAMVIERILRQVKINDYGTPTPTMAKHAKEIPN"
 9673..9989
 /note="5' LTR"
 join(10528..12588,12673..13083,13175..13473,13550..13808,
 13901..14083,14241..14306,14743..15045,15216..15479)
 /note="EST C28952(C62945) corresponds to a region of the
 predicted gene.
 Similar to maize transposon MuDR mudra protein isolog
 (AC003981)"
 /codon_start=1
 /protein_id="BAA92398.1"
 /db_xref="GI:7228438"

/translation="MTSKGIDRPAERTQSIYSWLLRGERIDOEYVYTWMSVWVYSRAT
 EGYFELMPLMDSTAARRVYVMAFERSPLVIFVSQEKDINVSQEDVSGPINSRAT
 VVPGSMOENEPREEQAMGMADEGERVGLVDENERSDNEERADDDSDDEGDMV
 ATDANEDFSLVILEGHDVFWYKENEVIGDARHTAHKDEKMEKVAHWASVQREVRV
 VKSTNIVTEVRCMKEDCPWRVAYKGNWYKVSIVTEHKYILOGVEKYHRTISAF
 VASEMYSSVGNIGPEPKSIIRHLENKFTYSYKACRAKQKITEMRYGTFEASYN
 LPRLLATARNNTYDILHTFTSYDDRTKSLQRAFTSLGACINAFVHCPRVLCIDG
 TFMGTGKRGQILTAIGCDGNQVVPMAFAFVSESTSWYFLEVRVHIAVVMRPNVC
 LIHHRHAGLRAIDYLONGWDEKRLPAKMPDVRSRMCMHMGANFYKQFKNHLMELF
 KRLCAKNOEKFNELWDLDELTKQDEQSRPQVEGDEPPIPLGALHDDPPTMRRR
 SGSSLRNFOTKTENEPTEKWSLLEDFTDTSRYGIMTNIAEYVNMVGRVRLPLVAIV
 EPIHLGTQAYFRDRYKKIGPSMADNNIVFGFEQVWRPLHLGPANTLPPTVSGSKAW
 LIQFTADLLHPDADNNRSVLEACTYLLWFGWVMTSTHGHVDFRVLHVARSTVDAQ
 PQDVPQWSGSAVLAATYRALCEATKTDAGAIAGQPMILQLMAARFAIRGPDVS
 APYGVGSAQWPEDEGPTMGTYCWRGRYAHVQVRGYPDPFVDFRLOPSDVTWEPY
 TEEAARAPLGLSSLCTRDQAYLTLLPVDFLVEPHCPQRYWRQFGLRQVPGWY
 QPTVLPADHSLTRQOLAGALWAPVQVYDDWVLATEEVINELFPHTNRYDLRW
 YLPTRTARVTFPDPAEPHVAATDAIPTHRDYFVGDADAARDISADITAVQYRLNR
 GLHTDVEORVTFDPMQEKMRVMRVSCHRSADVIVPPAGPDVLCRLAPLASQEDIQ
 HRFRTTYRPMCNFTIEWDMENPQNDGTRAYPRSETRSDYLRKRKDEHCECIAALEWH

VNPLGLPTWRPERECRGRCHVIRSVRHRTGQRCFCVCPNIVDDDFVFIHAWMEYER
 RVEIASLWNNPLGLPKWSEPKCGCHDRMQVATSFQAEQFRRRIYFVCPNVDNDFVV
 CLTIYKIRYDLHKITQTE"
 LTR 16155..16469
 /note="3' LTR"
 CDS join(23418..23748,25558..25589,25715..26480,26838..26965)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92399.1"
 /db_xref="GI:7228439"
 /translation="MGANCIATAKERPOCVPIEVSAFNRNVRHSPSFRWDRNTHI
 EDIMWPAFNSHSSGSIIRPETHSGSIAPDTGDFSGSPSDMENKIKCHKSDRKHSS
 KIADSGRLDSDNSSEAKLSRSLDTVSASDSKWSISVSPSTPAISRADPSSS
 RGHSLPTDASMRKARSPGYQYIRQVSDSKIPISLKSNEGASPPGESSMLSVCSN
 DLSVSGHSGSDGSMRTSEMAVSSQREWSVDSELLGVSVMKMTKSNANPNPTH
 SPQVECKLCKLKERSTWNAQELAVVALLCGHYHADCLDSLTADRYDPCPY
 CTHGEKTVKLFKLESTKTKPNKIVDVNLDGSSKHQEKIPLLEELYSQARSM
 RELKSRIGANDMTWLLNDLRIGPS"
 CDS complement(join(27953..28624,29565..29918,30369..30646,
 31026..31191))
 /note="ESTs C96615(C10106),C26336(C12127),D21959(C10106)
 correspond to a region of the predicted gene.
 Similar to NAM (AL021889)"
 /codon_start=1
 /protein_id="BAA92400.1"
 /db_xref="GI:7228440"
 /translation="MESLRDMVLPPGFGPHKDTLISHLYKKKHGQKIEVEIIEPV
 DIYKHPWDLPAKCDPTQDNKWHFAARDKRYPNGRSNRATVAGYWKSTGDKRAIK
 MGKOTIGTKTIVFHEGPRPTGRTETIMHEYIDEREQACPMKDAVLCRITKRN
 DWTPGNGELNDSHPEDYDAPPSVISTEQLNPAEPVVGVEAAPVTVAEPDGTTRS
 AITANIPSSDDILNDLWELNEDFPDFPEQSLASADLSPDEONVSSNVALAPKVE
 QDYSNENNVDDTEYLLPEDVINLHPGTDDFNMLQNLDDQTPQYATDVSNGTQKE
 ELWSPQANAEPSQSEADNDGIIRYRSMTKPTSPQPKGTQAKMRGVINGMATIS
 SESINOTIKFENSGRIVEHQKQADHASTKRSACKGKPTLSSNRGFLRGIRNAPAG
 CSDARNMILVAGFAIGVAVVALHIGQRLGLSQROQHT"
 CDS complement(38747..38998)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92401.1"
 /db_xref="GI:7228441"
 /translation="MMHGLVIRADTARPONMWAAPGPEVRHVRHGTPEVPVLCRPD
 DRGCRASLMMAMRLGLELVHFGCCYDDPEHPGPGCTSI"
 CDS complement(39315..39726)
 /note="3' LTR"
 CDS complement(join(40209..40396,40917..43545))
 /note="EST C28952(C62945) corresponds to a region of the
 predicted gene.
 Similar to maize transposon MuDR mudra protein isolog.
 (AC003981)"
 /codon_start=1
 /protein_id="BAA92402.1"
 /db_xref="GI:7228442"
 /translation="MSSKVTQFVHGEGRNIRFGPDGVDLSDFVMTSGIDRPAERTFO
 SIYSWLLRGFRIDQEVYTVMSVVSVRATEGYFELWMLPMDSTDAWRVYEMAFESWSP
 LTVFSVOERDINVSQTEDEGP INAGDVVGPSSMOENQPREEOAGMADGERVG
 IIVDEMERSDNEVEDDASSDEEGDYMATDNEDFSGLVISEGDHVPWEYKENEV
 IEGARYAHKDEKAVHAWVASLQREFRVVKSNIVYEVRCMKEDCPWRVHAKGKN
 DYKSVSTVTEHKYQCYGVEYRHNTISAFMASEMYSVVGNIQFEPKSIIRHLENKFK
 YTISAKAWRAKQKIIEMRYGTETEASYNLPLRIIAIARNNNTYYDLHTTYSDDRT
 KSVLQRAFSLGACINAFVHCRPVLICIDGTFMGYRGQIITAIQNGNQLPMAAL
 FVSENTESYWFLEVRHIAVVRMRPNVCLIHDRHAGMLRAIDYLONGWDEKGLPAK
 PDVRSWCMHGMANFYKPKNHILMELPKLCAQNECKKENELWDKLDLTTKOTDE
 QSRPLVEGDEPPIPLGALHDDPTMRRSGSSIRNFTQWENEPKPKWLLFDGGS
 RYRIMTNLAIVNWRGVRLPLVAIVELFHQTQAFYFRDKIGPSMADNIVF
 GNVYTKYMDKIKKARHVRVAQGTQVHRVIMCVRSRKGIVYKQACVCLKADGG
 CTCSCMKPFLHLHPCSHVLAAGDGPVPPNVYSNRYERKAIPTWSEETVYGFISGS
 YTTLSAOFVYLPDPSKLRVKKRRGRTIRINDMDESEAGRTLCRCKDLRGHTYKRC
 PKNAEVPSGADASPQGSADVLEFTHWLEQLMLTTTHFRCTTVVSDPVARVHNSTHE
 KRQELSLSTFGAAYLHKVHAHA"
 CDS complement(43999..44385)
 /note="5' LTR"
 CDS join(44572..44723,44846..44938,46366..46378)
 /note="hypothetical protein"

/codon_start=1
 /protein_id="BAA92403.1"
 /db_xref="GI:7228443"
 /translation="MALPRGVRSPKPMAAAAASASPPQPRGHMLVLTGTFGVGRYYSQ
 RLIAQGRVSGTSTSPAKKTELEMLGMDSTFADTSRCVM"
 CDS complement(join(48119..48257,48429..48703,49975..50139))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92404.1"
 /db_xref="GI:7228444"
 /translation="MALVNDKSDDDVNPACARRAAPLVIGAWGVFLQIARLGECE
 EPDRFRARKMAIISFCALSLLSFKNTRIYVLLCYLRKNYRPLIVTESKRTKCCRIT
 ASIARHGDELVCICLARLASSALDRVFKILKTDFEVLKMKRICVICINPVHISRTH
 GSSMYSTLPDAPVPRSTERLAGGIVICTTGG"
 CDS complement(join(51918..52637,52684..53346))
 /note="Similar to Arabidopsis thaliana chromosome 1 YAC
 YUPH12R sequence, C2-HC type zinc finger protein
 C.e-Mytl. (AC002986)"
 /codon_start=1
 /protein_id="BAA92405.1"
 /db_xref="GI:7228445"
 /translation="MASCINGELHVRVHRLVKASDESIRPHVLAVSNLDLIPRTI
 QVSMFCIYPKPTGTDGHDVVAFAAGLPSLLNHFPLAGRIVSNPCSGLPETIHCNQ
 GAELVVGADVALASLDYGTGVSAGVKILLPYAGDVALSVQVVSFAGGFTVAGTTH
 VVVDGSLMLVSAWSSELARSGTILAAAGARNHDSRVSFRSPSPSGASLDRAFTPLDG
 ARQADJARLREQARTRVOAVSAYLWKALAAVVSRRDARCMVWVMDGRRRLTSSS
 PELRAMRSYGVNTTFAVATMEETQKPLAEVASMARDATAAPAGGEHFOELVOW
 VEEHAGKORTYIDTASVGLGSPFASVTFASFEDTDRGHGHAAMALPTSSSARLCT
 FGQVTAARPGGDSWIASALLMPLAAALSDERKLGRIFRFTVETAYELGLRPSGSSA
 KRAGVITTSKI"
 CDS complement(57570..58009)
 /note="3' LTR"
 CDS complement(join(58018..58891,58980..62098))
 /note="Similar to Arabidopsis thaliana chromosome II BAC
 F9013 genomic sequence, putative retroelement pol
 polyprotein. (AC006248)"
 /codon_start=1
 /protein_id="BAA92406.1"
 /db_xref="GI:7228446"
 /translation="MRVNLQGLWTATDPGVAEPFEDMAALSAILOAVPREMLRGLA
 KIDTAKAWDAIKTRWGVDRVREAKEGRQFESMFKSETEPEEFAMRLTAVVAD
 IIRDMGQDEHEDVHNKKLLRVKPKYVAISLEQLLDLVKTHALEELVGRLSYVSD
 DEEGSDGKLYLTERQWQARVQKREGSGNGRGAPGONHGRKPGSPGKGA
 ATGANSRDISRVKFCNCFDGFYQACQKPRQRGRANLVQAAEEPTLLMAHVAF
 VSLAGEATLGRTPGGQEVHLTEKVLIDHEDGEGEEVTDWFLDGLTRNHTMGVRSFA
 AELDGVGVGVKFGDGSYIEIQGRGVTVVFRCKNGDHRSLDTVYIIPKLRKNLIISVRL

Query Match 64.0%; Score 16; DB 13; Length 146081;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catgccagcattccg 17
 |||||
 Db 8193 CATGCCAGCATTTCG 8178
 RESULT 30
 AC007476/c
 LOCUS AC007476 171705 bp DNA INV 17-MAR-2001
 DEFINITION Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
 BACR37P06, complete sequence.
 AC007476
 ACCESSION AC007476
 VERSION AC007476
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 171705)
 Cealiker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Branden,R.C.,
 Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,

Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Huck, J.J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattai, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, B., Scapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2R, region 49A-49B
Unpublished
2 (bases 1 to 171705)

Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazef, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R.R., Wan, K.H., Weinburg, F., Zhang, R., Zieran, L.L. and Rubin, G.M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2001 this sequence version replaced gi:5670614.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
source

1. 171705
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="49A-49B"
/clone="BACR37P06 (D594)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"
47654 a 38437 c 38157 g 47457 t

BASE COUNT
ORIGIN

Query Match 64.0%; Score 16; DB 4; Length 171705;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcatttcgatctcgt 25
|||||
Db 106124 GCATTTCGATCTCCT 106109

RESULT 31
AC037435
LOCUS AC037435 183937 bp DNA HTG 15-JAN-2001
DEFINITION Mus musculus chromosome 11 clone RP23-98I3 map 11, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
AC037435
ACCESSION AC037435
VERSION AC037435.2 GI:12229535
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1. (bases 1 to 183937)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus chromosome 11, clone RP23-98I3
Unpublished
2 (bases 1 to 183937)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bonkhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 2001 this sequence version replaced gi:7528163.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8882

Center clone name: 98_1_3

----- Summary Statistics

Sequencing vector: M13; M7815; 4% of reads
Sequencing method: Plasmid; n/a; 96% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 170056 bases at least Q40
Consensus quality: 177644 bases at least Q30
Consensus quality: 180041 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 base.
NOTE: This is a 'working draft' sequence. It currently consists of 24 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. as soon as it is available and the accession number will be preserved.

1 23881: contig of 23881 bp in length
* 23882 23981: gap of 100 bp
* 23982 24469: contig of 488 bp in length
* 24470 24569: gap of 100 bp
* 24570 24755: contig of 186 bp in length
* 24756 24855: gap of 100 bp
* 24856 25997: contig of 1142 bp in length
* 25998 26097: gap of 100 bp
* 26098 26764: contig of 667 bp in length
* 26765 26864: gap of 100 bp
* 26865 28426: contig of 1562 bp in length
* 28427 28526: gap of 100 bp
* 28527 30488: contig of 1962 bp in length

us-09-396-196f-7.oli.rqe

[illegible]

Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 266133)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7303437.
 Location/Qualifiers
 1..266133
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 join(392..425,1007..1155,5710..6105)
 /gene="Cam"
 /product="CT41657"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 <392..>15192
 /gene="Cam"
 /note="CG8472"
 /map="48F5-48F5"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(405..425,1007..1155,5710..5884,10160..10402,14732..15192)
 /gene="Cam"
 /product="CT41655"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(596..619,1004..1155,5710..5884,10160..10402,14732..14876)
 /gene="Cam"
 /product="CT41659"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(1153..1155,5710..5922)
 /gene="Cam"
 /note="Cam gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303486"
 /translation="MADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGONP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 4]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58542.1"
 /db_xref="GI:7303486"
 /translation="MADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGONP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AALRHVMTNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"

/note="Cam gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58541.1"
 /db_xref="GI:7303485"
 /translation="MADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGONP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AALRHVMTNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(5527..5642,5710..5884,10160..10402,14732..15192)
 /gene="Cam"
 /product="CT24787"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(5640..5642,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 3]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58543.1"
 /db_xref="GI:7303487"
 /translation="MADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGONP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AALRHVMTNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(<21254..21464,21584..>21600)
 /gene="CG13166"
 /product="CT32407"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 <21254..>21600
 /gene="CG13166"
 /map="48F5-48F5"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 join(21254..21464,21584..21600)
 /gene="CG13166"
 /note="CG13166 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 /protein_id="AAF58540.1"
 /db_xref="GI:7303484"
 /translation="MSERGGGAVTGRHARVSTPGRGSKTAYTPCRSEPPQDLAL
 AVPPPRNCIAIPMSLSLOTANGPGTKELHNP"
 join(<34383..34566,36219..36635,37304..37431,37529..37772,37837..37951,38022..38163,38232..38305,38481..>38679)
 /gene="CG13165"
 /product="CT32406"
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 <34383..>38679
 /gene="CG13165"
 /map="48F5-48F5"
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 join(34383..34566,36219..36635,37304..37431,37529..37772,37837..37951,38022..38163,38232..38305,38481..38679)
 /gene="CG13165"
 /note="CG13165 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 /protein_id="AAF58539.1"
 /db_xref="GI:7303483"
 /translation="MFRKNFRGASKSGGGGKLGSKANRTKSRDMLGVQOPREIHY
 RTHLFFSPNRPGYDVGECRSALSGIPSTPLSSTAPSLALPCLNLPFLDMPQPLVDRR


```

ENATDKCWRCDPNAENRQKMACALGFGSNAIGKGLRIYVVTNSDDVDVDPKPGT
LRYGVIOKEPLWIFGKNMKIKLSRELIVTSNKTIDRGFNHVIQNCAGIKIQASNI
IISNRJHNIVPTPGGLRESDEHVLGRSGDEGDGISIFSSHDIWIDHISMSRATDGL
IDAVAASNTNITISNCHFTDHEKVMFGANDHYVLDKDKMTLAYNHFGKRLDQMRPC
RFGEFHLVNDYTHWERYAUGGSGGATIIISQGNRFIAEDELLVKEVITYREKLTASV
WMKWTVISDGDDEMGNGATFTPSGDQNLDDKIDHNLIKPEPSSKVGILTFSGALSCV
KGRPC"
mat_peptide      546 a 338 c 353 g 476 t
/product="pectate lyase"
intron           902..999
/number=1
exon             1000..1300
/number=2
intron           1301..1397
/number=2
exon             1398..1713
/number=3
BASE COUNT      546 a 338 c 353 g 476 t
ORIGIN
Query Match      60.0%; Score 15; DB 14; Length 1713;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 qcatcgccagcattt 15
|||||
Db 486 GCATCGCCAGCATTT 472

RESULT 35
HSSVNTRO5/c
LOCUS          1907 bp mRNA PRI 27-MAY-2000
DEFINITION    Homo sapiens mRNA for syn trophin 5.
ACCESSION     AJ003029
VERSION       AJ003029.1 GI:8247276
KEYWORDS      syn trophin.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     1 (bases 1 to 1907)
AUTHORS       Piluso,G., Belsito,A., Puca,G.A. and Nigro,V.
TITLE         Identification of a novel syn trophin-like cDNA
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 1907)
AUTHORS       Nigro,V.
TITLE         Direct Submission
JOURNAL       Submitted (11-NOV-1997) Nigro V., Istituto di Patologia Generale ed
Oncologia, Seconda Università degli Studi di Napoli, Larghetto S.
Antello a Caponapoli 2, 80138, ITALY
FEATURES      Location/Qualifiers
1..1907
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="p25"
132..1751
/codon_start=1
/product="syn trophin 5"
/protein_id="CAB92969.1"
/db_xref="GI:824727"
/translation="MGTEGPPPPAASRGROGCLLVPAKTKTIALLYDESENAVDIR
LKLKRLVLTOKQDVCCVGGSHOGRNRRTVTLRRQPVGGGLSLIKGSEHNVPVVISK
IFEDQADQTMGLFVGDVAVIQVINGIHVENATHEEVHLLRNAGDEVITVYELREAPA
FLKPLGSPGPGSDSHSGASSPLFDSGLHLNGNSVTAPSSPSSPTAKDPYRKRWLD
TSLVPLSNARISRYKAGTEKLRNAFEVLALDGVSSGILLFYTAQDGTDLRAVSANI
RELTLQNMWANKCCSPDQVVMGWVNEKLQASDSQTRPRKFLALKGFSFVFSIP
PVSTFDVRAERYHLCVLEVKVKKFWLHEDCLQANLYLGLQDFDFEDQRPVCFSTP
AGHGSYFVNFVGLSELAWKESQFQATFMEVQTRVGYTMCWQSGEMLCFTVDFALG
FTCESKTKNVMRFKFSQKSSDDGKTRVKLLFNQLDTKQIEMKELFQDLRAVLH
CIHSFIAAKVASVDPPGPMDSQSLARKMYSS"
BASE COUNT    474 a 476 c 537 g 420 t

```

ORIGIN

```

Query Match      60.0%; Score 15; DB 93; Length 1907;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16
|||||
Db 570 CATCGCCAGCATTT 556

```

RESULT 36

```

AB048915/c
LOCUS          2616 bp mRNA PRI 20-SEP-2000
DEFINITION    Macaca fascicularis brain cDNA, clone:QnpA-17839.
ACCESSION     AB048915
VERSION       AB048915.1 GI:10241989
KEYWORDS      fis (full insert sequence).
SOURCE        Macaca fascicularis adult male brain parietal lobe cDNA to mRNA,
clone_lib:macaque brain cDNA library QnpA clone:QnpA-17839.

```

ORGANISM

```

Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (sites)
Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished (2000)
2 (bases 1 to 2616)
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
Direct Submission
Submitted (14-SEP-2000) to the DDBJ/EMBL/GenBank databases.
Katsuyuki Hashimoto, National Institute of Infectious Diseases,
Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku,
Tokyo 162-8640, Japan (E-mail:khashi@nih.go.jp,
URL:http://www.nih.go.jp/yoken/genbank/,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACCATGTGTG)
R. Site2: DraIII (CACCATGTGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al.(University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAAGCTGCG]:
3' end primer [CGACCTGCAGCTCGACACA] ).

```

FEATURES

```

Location/Qualifiers
1..2616
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QnpA-17839"
/clone_lib="macaque brain cDNA library QnpA"
/dev_stage="adult"
/sex="male"
/tissue_type="brain parietal lobe"

```

BASE COUNT

```

739 a 527 c 572 g 778 t
ORIGIN

```

Query Match

```

Best Local Similarity 100.0%; Score 15; DB 85; Length 2616;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16

```

```

Db 1743 CATGCCAGCATTTTC 1729
|||||
PSEIAAMH 3677 bp DNA BCT 26-APR-1993
LOCUS syringae tryptophan 2-monoxygenase (iaam) and indoleacetamide
DEFINITION hydrolase (iaah) genes, complete cds.
M1035 M35690
M1035.1 GI:151289
KEYWORDS indoleacetamide hydrolase; indoleacetic acid; tryptophan
2-monoxygenase; tryptophan-2-oxyreductase.
SOURCE P.syringae savastanoi (strain EW2009) DNA, clones pLUC2, pCP3, and
pCJP12.
ORGANISM Pseudomonas syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 202 to 3677)
AUTHORS Yamada,T., Palm,C.J., Brooks,B. and Kosuge,T.
TITLE Nucleotide sequences of the Pseudomonas savastanoi indoleacetic
acid genes show homology with agrobacterium tumefaciens T-DNA
Proc. Natl. Acad. Sci. U.S.A. 82, 6522-6526 (1985)
JOURNAL 2 (bases 1 to 312)
AUTHORS Gaffney,T.D., da Costa e Silva,O., Yamada,T. and Kosuge,T.
TITLE Indoleacetic acid operon of Pseudomonas syringae subsp. savastanoi:
Transcription analysis and promoter identification
J. Bacteriol. 172, 5593-5601 (1990)
JOURNAL 91008926
MEDLINE The iaah and iaam genes are part of an operon that is borne on a
COMMENT plasmid, pIAA, in oleander strains of the pathogen. There appears
to be no genetic transformation of host tissue by P.savastanoi. No
Pribnow boxes were found, though potential ribosome binding sites
are located at positions 435-439 and 2164-2168. Expression of iaah
depends upon the promoter for iaam.
Draft entry and computer-readable sequence for [1] kindly submitted
by T.Yamada, 14-JAN-1986.
Draft entry and computer-readable sequence for [J. Bacteriol.
(1990) In press] kindly submitted
by T.D.Gaffney, 28-JUN-1990.
FEATURES
Source Location/Qualifiers
1. 3677
/organism="Pseudomonas syringae"
/sub_species="savastanoi"
/db_xref="taxon:317"
19..47
/misc_feature
/feature="indoleacetic acid operon promoter"
54..3537
mRNA
/feature="indoleacetic acid mRNA"
82..312
CDS
/feature="promoter proximal ORF"
/codon_start=1
/transl_table=11
/protein_id="AAA25851.1"
/db_xref="GI:151290"
/translation="NPGRTPANWLDKAMVAVARIRQKPKQAVPAALLQAPHAMEFVW
STKRTFTECHCPAYTPYTQASAWSASAVVSAT"
454..2127
CDS
/feature="tryptophan 2-monoxygenase (EC 1.13.12.3)"
/codon_start=1
/transl_table=11
/protein_id="AAA25852.1"
/db_xref="GI:151291"
/translation="MYDHPNPSIDILYDYGFLKKCKMTGGIGSYSGTPTPRVAIV
GAGISLVATELLRAGVKDVLYESRDRIGRWVSQVFDQTRPRYIAEMGAMRPPS
ATGLHYLKKFGISTPTDPVGVVDTELHYKRWHPAGKKPPELFRVYEGWQSL
LSGVLLEGSLVAPUDITAMLSGRLEEAIAWQWLNVRDCSFYNAIVCIFGTRH
PFGDWRAPDFELSGSIGSGGFLPVQAGFTLLRWVINGYQSDORLPIDGISS
LAARLQSQFDGKALDRVCFRVRGRISREAEKIIQIAGEQRVDFRVIVTSSNRAM
QMTCLTDSFSLSRDVARVRETHLTSGSKLFIPLRTKFWINKLPPTIIQSDGLVRG
VYCLDYQDPEHGKGVLLSYSTWEDDAQKMLAMPDKKTRCQVLDLALAIHPFTFASYL
LPVDGYERYVLHDMWLTDPHSAGAFKLNYPGEDVYSQRLFFQPMPTANSPNKDITGLYL
AGCSCSFAGGWIEGAVQTALNSACAVLRSTGTGQLSKGNPLDCINASYR"

```

```

CDS
2170..3537
/feature="indoleacetamide hydrolase"
/codon_start=1
/transl_table=11
/protein_id="AAA25853.1"
/db_xref="GI:151292"
/translation="MHEITITLESICQALADGEIAAAELRERALDTEARLARCIFIRE
GDVAVQFGEADHAMKGTPLMGMPVSKDNICVRGLPTAGTRGMSGFVSDDAAIVSQ
LRALGAVAGKNNMHELSEFVYTSINPHWCTVGNVPVPCYAGSGSSGSAVASGIVP
LSVDTDTGKOSIRIPAAFCIGTFRPTTGNWSTAGLIPVSHTKDCVGLLRTAGDAGFL
YGLLSGKGSFPLSTAPCRIGLIPVMSMSDLDEVERACVNAISLLRKTGTGFEIIDD
ADIVELNOTLTLTVPVLYEFFADLAQSLLSLWKGTHIHFIAQVDDANVKGIIHHHLE
GALKPAHYLSLQNGELLKRMDELFAHNIELLYPTVPCRPVHLADHAREFFESQA
IINTDLSNAMLPSTIPVPEGRPLVGLSFDALRGDRALLSRVSAIEQVLGFRVRY
LPHTT"
BASE COUNT 813 a 917 c 1046 g 901 t
ORIGIN 1 bp upstream of EcoRI site.
Query Match 60.0%; Score 15; DB 3; Length 3677;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 qcctgcgcagcattt 15
|||||
Db 1729 GCATCGCCAGCATTT 1715
|||||
RESULT 38
AB028980/c 5618 bp mRNA PRI 04-AUG-1999
LOCUS Homo sapiens mRNA for KIAA1057 protein, partial cds.
DEFINITION AB028980
ACCESSION AB028980.1 GI:5689450
VERSION Homo sapiens brain cDNA to mRNA, clone_lib:pbuescriptII SK plus
KEYWORDS clone:hhl1838.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hirose,M., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XIV. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 6 (3), 197-205 (1999)
JOURNAL 99397452
MEDLINE 2 (bases 1 to 5618)
REFERENCE Ohara,O., Nagase,T. and Kikuno,R.
AUTHORS Direct Submission
TITLE Submitted (17-JUN-1999) to the DDBJ/EMBL/GenBank databases. Osamu
JOURNAL Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
Source Location/Qualifiers
1. 5618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="hhl1838"
/clone_lib="pbuescriptII SK plus"
/tissue_type="brain"
1..2935
/gene="KIAA1057"
<1..2935
/gene="KIAA1057"
/codon_start=2
/product="KIAA1057 protein"
/protein_id="BAA83009.1"
/db_xref="GI:5689451"
/translation="LADSSPNLQIIIKELLSMHHPDPALTKKFDYLPVVDSSSSG

```


FVGLRNGGATCYMNAFOOLYMOPGLPESILLSDVDDTDNDPDSVFFVQVQSLFGLHMS
 KLOYVVPENFWKIFKMNKELYVREODAYEPFTSLIDOMDEYLKMKRQDIQKNTFQ
 GYSDOKICKDCPHRYERERAPMALNMGVTSQSLSEISLQDFVRGEVLEGSNAYICEK
 CKEKRIYVKTCIKSLFSVLVILHMRGFDWESGRSIXYDEQIRFPWMLNMEPYTVSG
 MARQDSSEVGENGRSDVQGGSGPRKKVALTENYELGVVHSGQAHAGHYYSFKD
 RRGCKGKWNVDVLEEDDLNDETLEYECFGGEYRPKYVDQNTYDVRVRRYWAY
 MLFYQVSDQNSVLPKRSVVRQEAEDLSAPSSPEISQSSPRPHRPNDRLS
 ILTKLVKGGKGLFVEKMPARVQVNDENLAKMKRDVYSSDYSEFVLSLASLAT
 KLRHPYPCMAKYSLOLAIOFLQTVLRTKKLRVDEWATIEALLSKSDACQWL
 VEYFISSEGRELIKILLECNRVREVRVATILEKTDLSALFYQDKLKLHOLLEVL
 ALLDKVPCNKCAQYFFLFNTFVQKQIRAGDILLRHSALRHMSFLLGASRONNQ
 IRWSSAQAREFNLHNTVALLVHSDVSSQNVAPGIFKQRPPIAPSSPLPLHE
 EVALLHMGKPYLLEVMPALRELTGSLIALTEVYVYCFQNEHESFTMLFIKNOL
 ETAPPHLKNCTFOLLHEILVIEDPIQAEVKEVFETENGLLALMHHSNHVDSRCYQ
 VKPLVLAQCPAAKVEYFKENSHHWSVAWQVLQKMKSEHYWTPQSNVNETSGTKFO
 RTSAQDTLAYATALLNEKQSSNGSESSPANENGDRHLOQGESPMGIELRSDL
 DDVDP"

BASE COUNT 1625 a 1103 c 1238 g 1652 t
 ORIGIN

Query Match 60.0%; Score 15; DB 85; Length 5618;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagatttc 16
 |||||
 Db 4770 CATGCCAGCATTTC 4756

RESULT 39
 AE004320/c
 LOCUS
 DEFINITION Vibrio cholerae chromosome I, section 228 of 251 of the complete chromosome.
 ACCESSION AE004320
 VERSION AE003852
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
 Sellers, P., McDonald, L., Uitterback, I., Fleischmann, R.D.,
 Nierman, W.C. and White, O.
 DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 cholerae
 Nature 406 (6795), 477-483 (2000)

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
 Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
 Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
 Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
 Direct Submission
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 10205
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:866"
 /chromosome="I"
 /note="biotype: El Tor"
 128. .1639
 /gene="VC2501"

TITLE
 JOURNAL
 FEATURES
 source
 complement(5205..6182)
 /gene="VC2504"
 /note="similar to GB:L42023 SP:p45250 PID:1007767
 PID:1221703 PID:1205791; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="2-hydroxyacid dehydrogenase family protein"
 /protein_id="AAF95646.1"
 /db_xref="GI:9657086"

CDS

128. .1639
 /gene="VC2501"
 /note="similar to SP:p11648 GB:X15130 PID:1054725
 PID:43309 PID:537102; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="aminopeptidase A/I"
 /protein_id="AAF95643.1"
 /db_xref="GI:9657083"
 /translation="MEFSVKSGSPKQPSACIVGVPEPRRLSPVAEOLDKISGYIS
 SLRLRGDELEGPCQMLLHVQPVLSERVLLVGGKRELGEROTKEIQIKIINTLE
 TGSWEAVCFLELHVKGKRDITWVKVQAVKATKGLYIPDFKSVKPEIRRLKLVFN
 VTRRELINERAITHGLAISGVKACKDLGNMPPNTANPAYLASQARRLADDESIT
 TKIIICEEMERKGLMASLAVAGSRNEMSMVIEYKGNPDPEAKPIVLVKGGLTFDSG
 GILKPGGEMDMKYDMCGAASVFTKATAKGLPLNVIGVLACGCMGPNAYRPG
 DILTMGQTVELNTDAEGRVLCMDVLTVVERFEPECVVDVATLTACVIALGHHIS
 AVMSNHPLAHELVNAEQQSDRAWRLPLADEVHEOLKSPFADMANITGGRRGGAITAA
 CFLSKFAKKYNWAHLDIAGTAKWSAAKSGTGRPVSLVLLVFLNRSGLDAEE"
 1723. .2172
 /gene="VC2502"
 1723. .2172
 /gene="VC2502"
 /note="similar to SP:p28905 SP:p11649 GB:L04574 GB:X15130
 GB:214155; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="DNA polymerase III, chi subunit"
 /protein_id="AAF95644.1"
 /db_xref="GI:9657084"
 /translation="MPTATFVLSRDSAAQATPEGLAQYVLYLTRHFAAGCARVLOCO
 DKHAEOLAEFRWOIEPEQFLAHNLGEGPKSGTALGIVGVGSPSNRQLAINLADN
 QTTTFARTFTEVDVPCNAENAKQLARERYKLYRQAGYQLOTVELIQP"
 2266. .5127
 /gene="VC2503"
 2266. .5127
 /gene="VC2503"
 /note="similar to SP:p07118 GB:J03497 GB:X05891 PID:147906
 PID:43302; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="valyl-tRNA synthetase"
 /protein_id="AAF95645.1"
 /db_xref="GI:9657085"
 /translation="MEKTYNPTIEQDLKYTWEEQGYFKPHGDTSKDYSIMIPPNV
 TGLSHMFADQDTIMDTLIRQRMKNTLWVGTDHAGIATOMVVERKIAEEGKTK
 HDYGRDAFIDKIWEKAEAGGTITKOLRLGASVDWDRERFTMDQGFYKAVQEVFRL
 YKDDLIVRGKRLVNDPKLHTALSDLLEVENKETKGMHFRPLADGVKTAGDKTIV
 VATTREPTMLGDTGVAVNPEDPRYKDLIKGKILLPIVRRIPVIGDEHADMEKGTGV
 KITPAHFDNDYEVGKRHLQPLMILTFDANIRDAEAVFNSNGEASNAYTEIPAKYQG
 MERPAAKAIVAEFEELGLLOETIKDHLDTVPYDGGVYIEPMLTDQWTVRAGILAKP
 AVEAVENGDTQFVKQYENMYFSWMRDIQDWGISROLWGHRTIPAWYDQGNVFCVRN
 EEEVRAENITADVALRDQDDVLTWTFSSALWTFGLGWPEKTPELKVPHPDVLVTG
 FDIITFVARMIMMTWHFCKDEGKAQVPKTVYVLTGLIRDENGKMSKGNVDPI
 DMIDGTDLESVAKPTGNMQQLAAIKENKTKTFENGIEAGYDLSRFTLAAMAST
 GRDINDMKRLGYRNFCKNLWNASRYVLMNTEEQDQCGFAAGAEYLSLADKWIESQF
 ELAAEFNGHIDFRDMAANTLYEPIWFOCDWLELTPLVLMKGTGTAQQRATRTL
 ITVLEKTLAHPVITPYITETIWSVKPLVDVEGDTIMQLQPYDVANFNOEALDD
 IEWKAFTISIRNRAEYINPKPLEVMAKAEQDAARIEANKPVLVSLAKLESIR
 VLADGATPACATALVKGSLMIPNAGLIDKDAELDRLAKEITAKTQGETARTIEGLGN
 EGFVAKAPEAVITKEREKLAGYQEAIVKLEQQKATIAAL"
 complement(5205..6182)
 /gene="VC2504"
 complement(5205..6182)
 /gene="VC2504"
 /note="similar to GB:L42023 SP:p45250 PID:1007767
 PID:1221703 PID:1205791; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11

gene

CDS

gene

CDS

gene

CDS

gene

/translation="MSLPTSLPTVFLDRATIPRIHISLPALPFEHHWLEYDACEPQQ
VVERLIAADIVITNREMLIQLPKLIAISATGTNNVDLPACRDLNIAVCNVQ
GYATRSVPHVAMFALRNLIAGYHNDIAEGWQRHGFCTFTPIGIDIASTMGIT
GSGALGOATANARALGHVLLAERKGOVECRDGYTSFEOVLAQSDVLSLHCPITDET
RNISAEALQNPALLINTGRGLVDEALVKRQTAGAGVDVFSAPADMND
PLIANDLPLNLLTHVWAGSSDQQLATILINDISAFMRGEAKRVV"
complement(6222..6959)
/gene="VC2505"
complement(6222..6959)
/note="VC2505"
/note="similar to SP:P39219 GB:U00096 PID:1786244;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ribosomal large subunit pseudouridine synthase
A"

gene

CDS

/protein_id="AAF95647.1"
/db_xref="GI:9657087"
/translation="MAETYNPKDPWDIVYEDHIIAANKPSGLSVGPRVVEHYD
SMWRLVDCPDVVKVHRLDMSTGLIFAKGHMSALKKGFQRLTHKIIYARWG
VMPQNEGEVNLPLCCDMPNPKQKVCFEGRDLSLTRYQVKKREKTTVVKLIYITGRS
HOURVHMSLGHPIVGDGLYAPPEAVEFADRLQHAALCFYHPRSHWLRTLFVPCDF
YPAEQIQLHDPDKLPDYKLSQD"
complement(7056..9965)
/gene="VC2506"
complement(7056..9965)
/note="VC2506"
/note="similar to GB:U00096 PID:1786245; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="RNA polymerase-associated protein HepA"
/protein_id="AAF95648.1"
/db_xref="GI:9657088"
/translation="MSPALGOWISDTSGLGTGVVVALDARTVTILMFAASENRVYA
RSDAPTRVFNQGVDDVDSQGSLOVEQGVVYLTGTRVDEESGVALREIFL
SNQIRFNKPODKFAGQIDRDNFVLYRALTQYQHKSPMGLCGMAGLIPHOLY
IAHVRGRRHAPVLLADEVGLKTIAGMIHQVLIQGRALRILVVPETLQHWLVE
MMRNFHFIIFDEERCEAFSEADNPFQTVQVCSLDFLRKSRFQALAEANDL
LVVDHAEHLHWPKEPQREYVIEALAEQTPGVLLTATPEQIGRESHFARLLDAD
RFYDEYFAVEEQYAPVADAVTALFSGEKLSEAKNKITELISEQDVEPLFKALSS
ASEDEIALAQELIDNMDRGTGRVLTFRATRAIKGFPVRNVHLLPLETSPQYTTSM
RVAGMLGKLTPEARANKMLYPEIIFQEEGESSWQFDSRVNLLKVKAKRSEKI
LVIASASTALQALREREGRATVFEHGMSTIIEKDAAYFAQEGGAOVLICSE
IGSEGNFQALQVLMFDLPNLDLQRIQGLRIGCKRDRIDVYVPLTETSQLILA
RWFQEGNLAETCPTGRAVYDAERLIPILAGGGELEVIIEESAKLNTKLSQL
EVGRDLLEHNSHGREQAQIAEQIAKTGDNTLVTFALSFDALIGHQEDGENALV
VTPAHHMVPSPGLPYEGATITPDRTALSRDMHFIISWHPMWQGGIDLLMSEGV
TCASVLLKNALPGVTILLEVYVVDQAAPKRGISRLFPVPIRILMDARGNDLSQ
VERESFNRLSPVNRHLASKLVSSVQDHLRLITASETAVEPRVSAIREQAQRDMQGS
LNSLERLALUKAVNPNDREIEIVLQOIKELTGITYAQAOQLDLSRLIVVHN"

gene

CDS

BASE COUNT 2663 a 2634 c 2514 g 2394 t
ORIGIN
Query Match 60.0%; Score 15; DB 1; Length 10205;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cagcattcgatcct 22
|||||||
Db 6278 CAGCATTTCGATCCT 6264

RESULT 40
AE003860
LOCUS AE003860 10593 bp DNA BCT 06-JUL-2000
DEFINITION Xylella fastidiosa, section 6 of 229 of the complete genome.
ACCESSION AE003860 AE003849
VERSION AE003860.1 GI:9104830
KEYWORDS Xylella fastidiosa.
SOURCE Xylella fastidiosa
ORGANISM Xylella fastidiosa

Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
1 (bases 1 to 10593)
Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van
Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
Meldanis,J. and Setubal,J.C.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
Nature 406 (6792), 151-157 (2000)
20365717
2 (bases 1 to 10593)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvares,A.R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A.A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J. de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meldanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
Location/Qualifiers
1..10593
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(277..1356)
/gene="XF0061"
complement(277..1356)
/note="similar to SP:P07674 (percent identity: 28 %/query
alignment coverage: 64.9 %/subject alignment coverage:
65.1 %); identified by sequence similarity; putative; ORF
%); identified by sequence similarity; putative; ORF
identified using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="transcriptional repressor"
/protein_id="AAF82874.1"
/db_xref="GI:9104831"
/translation="MLLQTLFELDRATAMSLFCNRIKPTQONOLATVAFSGEITEE
IDPELIDLEKMRSKNPGFTIELMELGNDMRQDQHEPVLIRKSKPGRYLMVAG
ERRWLACAVDIKIKAVIRELTDQVVRVRAENIQRENIQLEIAVALRDKERLGT
LEKVAENWKINNVSRINFLVETNSLQASVSEGTADITANDIRLEKVNQT
AAKTIIDQAKVDPINRKTVREKLQNAKPTTNEKNATDSRQESQSTFOEGETHLQ
NISNOEKPAVASAPLPDPAQALVALAAKQMETLCTEVSALSLSKILHQRDVIALGDV
LQNVVQAALFTRHVLASLANSK"
complement(1331..1657)
/gene="XF0062"
complement(1331..1657)
/note="hypothetical protein; identified by sequence

```

similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82875.1"
/db_xref="GI:9104832"
/transl_table="MYHRSIFORSNSIHNETVHKWSDLQNHDIOTYPLPSTNNI
LEWQNFKNTEICAENTSPIQSYRPRPEQHVNFPIILPHLMPVQHLKMYQARHA
VWTDIV"
complement(2427..3152)
/gene="XF0063"
CDS
complement(2427..3152)
/gene="XF0063"
/notes="similar to GI:5921548 (percent identity: 37 %/query
alignment coverage: 93.4 %/subject alignment coverage:
95.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="competence protein F"
/protein_id="AAF82876.1"
/db_xref="GI:9104833"
/transl_table="MKDAVNSNHILPDTWCDRLLWLLPSYCLACNEPCHOCKTLCAT
COLLPRMENACLRCAAPLKELEGHTPCGCLYDPPLETAHAFLYHWPINKLISR
FKPHDLACGHLLADNMAAQAHPRPQALVPSLHHGLRTRGQDALELARHYGNR
LQPLCLPALRRVATLQPSQLNKRNSNLNAPFIHPPLPAHALVDVDDVMTGTTLH
AAARLHGTGVRIDAWICARVP"
3202..4227
/gene="XF0064"
CDS
3202..4227
/gene="XF0064"
/notes="similar to SPIP36569 (percent identity: 60 %/query
alignment coverage: 94.4 %/subject alignment coverage:
93.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: 66"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF82877.1"
/db_xref="GI:9104834"
/transl_table="MSSIRYDWTAEELHALFDLSPELRYAASVHRHQHFDPAEIOV
STLLSVKGGCEDSCYPOAGRYDTVTAQKLMDDVAVAKQAQKLAGASRFGCMGA
AWRSPKDRDIPKIAAMIREVRLAGLETATGLMLNTCOAQKAGLDYYNHNDTSP
DFYDSVHTROYQDRDLTAHVRDVLGKTCGCGIVMGETQHRHVRGLLLTLATLPAHP
DSVNVLLVQVATPLHCTOTLDPEFFVMTAVARITMPSRWRLSAGRESDELQOL
LCFMAGANSIYGEKLLITANPETERDQALFQRLGRPMHLMENVNQDQHGHVHAD
TACKHV"
4281..4382
/gene="XF0065"
CDS
4281..4382
/gene="XF0065"
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82878.1"
/db_xref="GI:9104835"
/transl_table="MATLVDVVSFWDGVNLDSDSGGERRKVVFEFS"
4486..5673
/gene="XF0066"
CDS
4486..5673
/gene="XF0066"
/notes="similar to SPI034731 (percent identity: 33 %/query
alignment coverage: 63.3 %/subject alignment coverage:
96.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF82879.1"
/db_xref="GI:9104836"
/transl_table="MSLINHKNGVSCNGSGTCSRDLNRLMLRRVVGNCNRPVPHLSY

```

```

LALSSMCTFRCSFIATIVWLLTACSSAPKSRSLSSVLSQSLSDISLPLRTGIALGG
GAAGFAHIGVLKMLEANGLAPSVVAGTSGSVVAGLYASGMDSKFQIOTKAVAMDESIN
IRDLFFSGGLVQGOQLQNVNELVGRPIEKLKPKGAIATRILEDGORTWFRGNYG
EAVRASCSIPGVFEPTIGAYHYDGGITSPVVDVAVQLGADFIADVISKATGKN
PGGLVGTNQSISIMGOHLGEAEUKRANIVIPKVLIDIGSTDFDORNAAILLEGGRATI
AVIPERLKLQALQOORIQAAKRTANQSAIEAVESYRRCQNSRTRWEKLRGKEDCTK
SSL"
6138..6377
/gene="XF0067"
CDS
6138..6377
/gene="XF0067"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82880.1"
/db_xref="GI:9104837"
/transl_table="MSCKEIFEYIVIKQWLPMEFFAOMGMVMGAINCADRYPPVILIS
RCVLSYVGMGYRVRFVSIVLPMCMYGLGVMIVLAF"
complement(6494..7495)
/gene="XF0068"
CDS
complement(6494..7495)
/gene="XF0068"
/notes="similar to SPIP26601 (percent identity: 52 %/query
alignment coverage: 86.8 %/subject alignment coverage:
99.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hydroxybenzoate octaprenyltransferase"
/protein_id="AAF82881.1"
/db_xref="GI:9104838"
/transl_table="MILNQEHQIYVNHILNSFQPLPNSDSSNPSVMAYERFTSAI
TLMLHMRNLDIPYKLRGDRPVGFLLMLTWMALAAAGVPPMWTLCVFTGIWL
TRSGACVINDYTRDLDPHVERTIRPLVTGAVSPRNALIMEATLMLIAPGLVLTWNO
LTVLLSVGLFLAMTYPLKRYTYLPQVYLGAFLGWPIMAFAGIOGKVPQAWLLYV
ANIITWYDTWCAAMVDRDDIKMGAKSTAILFADLDLTVOGVLYTLMLFTLCLVGLR
ATLSHYWISLIAAVALIGYQFIARREPTACFRAPMHNWVGMIFAGIALATTH"
7506..7582
/gene="trna-Arg-1"
/notes="found by tRNAscan"
/product="trna-Arg"
7506..7582
/gene="trna-Arg-1"
complement(7517..7909)
/gene="XF0069"

Query Match 60.0%; Score 15; DB 1; Length 10593;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgactct 22
|||||
Db 3304 CAGCATTTCGATCCT 3318

RESULT 41
AE004795
LOCUS
DEFINITION
Pseudomonas aeruginosa PA01, section 356 of 529 of the complete
genome.
ACCESSION AE004795 AE004091
VERSION AE004795.1 GI:9949931
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 10978)
Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Wartner,P.,
Hickey,M.J., Bramkan,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,

```

Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
 Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.
 Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen
 Nature 406 (6799), 959-964 (2000)
 20437337
 2 (bases 1 to 10978)
 Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
 Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
 Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
 Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
 Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H.,
 Hancock, R.E.W., Lory, S., and Olson, M.V.
 Direct Submission
 Submitted (16-MAY-2000) Department of Medicine and Genetics,
 University of Washington Genome Center, University Of Washington,
 Box 352145, Seattle, WA 98195, USA
 Location/Qualifiers
 1. 10978
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"
 complement(132..449)
 /gene="PA3762"
 complement(132..449)
 /gene="PA3762"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG07149.1"
 /db_xref="GI:9949932"
 /translation="MYKLCFYVPESHLDVVKQAVFAAGGGRIGAYDSCCWSLQGGQF
 RPLDGSQPYLGQGVQVHEAEMKVELVVADELHSHVSKALKAHPYETPAYEVRWLTD
 RVF"
 complement(914..4810)
 /gene="PA3763"
 complement(914..4810)
 /gene="PA3763"
 /codon_start=1
 /transl_table=11
 /product="phosphoribosylformylglycinamide synthase"
 /protein_id="AAG07150.1"
 /db_xref="GI:9949933"
 /translation="MLILRGAPALSAFRHGLLEQLTOHVPAVTGLYAEFAHFADVTG
 ALTADAEQVLAFLKYPSPVQEPGRFLVLPFRFGTISPWSKASDIARNCGLAKI
 DRLERGIAYVQGELESQDAQVAARLHDMRTQLVLDLEGAELFSAQPRPLTAVD
 VLGGRAALEANVELGLALAEDEIDYLLKSGELGRNPHDVELMFAQSEHCRHK
 IFNASWDIDQADKSLFGMLKNTYEMNREGVLSAYKDAAVIVGHVAGRFPPDQTR
 EYASREPQVILMKVETHNHPTAIAFFPGASTGSGEIRDEGATCRGAKPAGLTGT
 VNSLOIPGEQPEWYQPKERIVTALDIMEVGLPGGAAFPNNEGRPALTYFRFEQ
 KIATPHGEVRYHKPIMLAGGMNIRDEHVKQGEISVGAKLILVGGPAMLILGGGA
 ASSMATGASDLDFASVORDNPERECQEVIDRCWLGERNPISFTHDVAGGLSN
 ALPELINDGRCRELRVPNDPEKMPLEIWCNESQERYVLSVDAADFETFKAICE
 RERCFAVGEAIEORLTVDNSHFDNKPVDMPLEVLGKAPRMHRAVTRAEALGDDF
 DAAGLEQSVRLRHPAVASKSLITIGDRTITGLVARDQMGPMOVPAVDCAVTA
 TSFVITGAMWAGERTPLALDDPASGEMALGETVYNLAARVGLSDIKLSNWA
 AAGHPGEDARLYDVKVGMELCPGLGITIPVGDNSMKTQWQNGEDKSVTSFVSL
 IYTGAPADVRSQTLTQQLDKGETDILIDLGRKNRGLSGSILAQVHGLGRAPVD
 VDDAELFAFFAIOQLNADGHILAYHDSRDLGLTSVLEMAFAGHCGVELNLDAID
 SREELAAVISELGAIVQREGATPEVLAQFSAAGLDDCAVIGQPVNGYEINLYN
 GETVYSNORILQITWSETSYOIRLONADCAEAFDALLDENPGLSILKLSYDND
 DTAATYIKKVRPKVAILREQVNGQVEMAAAFDRAGFAADVHMSDILACGVCQDMNS
 LGLVACGGSFYGNARVGGGWAKSILFNARVAMVQVQESSIFDGMAGSRPLPIAHGEGHA
 EFSEBALLEADLSCGSLRVFDVNHGKVTEAYPANPNSPRGITGLSSRDGRVTTMMP
 HPRVFRVQNSWRPDDWQEDGGWLMFRNARVVD"
 5211..6569
 /gene="PA3764"
 5211..6569
 /gene="PA3764"
 /codon_start=1

/transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAG07151.1"
 /db_xref="GI:9949934"
 /translation="MKEGVLRVITRNSPATYFQDRNGETFEYELAKRFAERLGLVEL
 KIETADNLDLYAOLSRREGPALAAAGLTTPGREDASVYSHTYLDVTPQIIRNGQQ
 LTRPEDLVGRKIMVLKSGSHAEOLAEKQYPELKYEFSDAVEVNVALLRWVDGID
 LTLVDSNELANNOVYFNVVRAVDFGEARGLAWALPGGDDSLMNEVNAFLDQAKKIG
 LQRLKDRYGHVDVILGYGATFAQHLQOBLPRYSEHKOSGKOLDTDRLLAIGY
 QESLWPGATSKTGYGRLMLNTRTAQAMGVSNRDLDPKQSKOSGSKYFVQISELPES
 IKEPDRSWFALAAYNIGAHLEADARKMAEKELGNLKWLDVKKMLPLRLAQKQWAKTR
 YYGAGGETVHEVQNVRRYYDILTWTQPMEGSQIAESGLHPLGVNKRTRPEDSGDE
 KI"
 complement(6583..7167)
 /gene="PA3765"
 complement(6583..7167)
 /gene="PA3765"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG07152.1"
 /db_xref="GI:9949935"
 /translation="MANDPYIQSFTCFDGTFRLLTAALPEVALALKQAVAGCAAGPLL
 VFDNATGRSVDLDIRGSSDILLARLVVAGERQPAEEAPRGGRFLGLVVADEVILLP
 RHWELAAQPGGASVALRKLVEARRSQSRQAQERAYHFPMNMAAGDLPGFEATR
 ALFAGDREGFAERIAGWPTDVRHAALASDGES"
 complement(7293..8549)
 /gene="PA3766"
 complement(7293..8549)
 /gene="PA3766"
 /codon_start=1
 /transl_table=11
 /product="probable aromatic amino acid transporter"
 /protein_id="AAG07153.1"
 /db_xref="GI:9949936"
 /translation="MSDTRVQVQDHAGAEAVDSSGLEVKRLTFLEAVAMIVGTNIGA
 GVLMSAYSRKAGFMPLLLWAVAGLFTTISMLIVSETALRTRTHQLSLGAAQVGS
 FGWAIFLSVAVNSIGALITVSGSGKILSAFFGISPALGSQVLFPIPAAGVILGLSA
 IGKGRKFSIGVSMILILVAATLLNDNTEFAFLDGDWITVYVFNIAVFCFSAQYI
 VPMARGFSAHRLPKAVITGMLTFFVLSIVPLSVIALTGLENSEVATLAWKAL
 GEWAFTANTFALCAMLTSYVGLGSGFLTNIFDKFKLGPENRKPTRLVLAIVAPP
 FVLASGLGVFNALYFAGAFSGVILSIMPILMLRSARRNGDFEPAWKCGWIAHPAIQ
 AVITYILASAAIATCSAMLLPAGW"
 complement(8938..9486)
 /gene="PA3767"
 complement(8938..9486)
 /gene="PA3767"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAG07154.1"
 /db_xref="GI:9949937"
 /translation="MRVTKLISNGKGYTVKVRPTIDRSRDPHFMRALAEAKAA
 LGEPVCAVLVREGEIIRGFNRPISSHDPSAHAEMLAIRMAAAEAGNYRLPGSTLYV
 LTFPCSMGSGLLVHARLQRLVYTVVEPKSCAVESGRFFEQEHLNHRVMVEGGVLAEE
 CSQALSACFFRAREAAKRPED"
 complement(9477..10868)
 /gene="PA3768"
 complement(9477..10868)
 /gene="PA3768"
 /codon_start=1
 /transl_table=11
 /product="probable metallo-oxidoreductase"
 /protein_id="AAG07155.1"
 /db_xref="GI:9949938"
 /translation="MTFTROVLGGLAVGLVAGGARLWLARPOVAQBYDYELTA
 APLDLIEVPGSSPALYGGCGVGRVRAKQGLVRLVRLNDEPTIHWGIRLPI
 EDMGVYISQPPVQGESFIYQFQDAGSYWYHPLMSSEQLGVLGVLIIEREP
 TCFRHEKVLCLKTHVDQGAFTFSPQAAREGTRGRYSTINGKIVPTIDLPAGOI
 VVRLNVDNTVTYRLNLPNGEARIIAIDHPVPRFGEGQVIGPGRNRLALKVPE
 AGTSLSRDGPVRLATRSVASEAPAGDWPKPLPANPSEPLDANAEKIGFREWVG
 AMSDTSGKNPYSPFWQINGKAWEGEGEHNAPLAKLKQGSYIFELRNAAQYOHPI
 HLHGAFVLDSDRDIIPYFTDITVLLCKNETARVALVDNPLGLMFMFCHVIDHMETG

```

BASE COUNT      1876 a 3658 c 3644 g 1800 t
ORIGIN
    LMGTIATGCAWCG"
Query Match      60.0%; Score 15; DB 1; Length 10978;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagcattcg 17
|||||
Db 9239 ATCGCAGCATTTCG 9253

RESULT 42
AC018333/c
LOCUS
DEFINITION      AC018333 23873 bp DNA HTG 09-DEC-1999
                  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
                  pieces.
ACCESSION      AC018333
VERSION        AC018333.1 GI:6552858
KEYWORDS       HTG; HTGS_PHASE2.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 23873)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10214399 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES             source
    source          1..23873
                    /organism="Drosophila melanogaster"
                    /db_xref="taxon:7227"
BASE COUNT      6912 a 4665 c 4568 g 7728 t
ORIGIN
    gcatcgccagcattt 15
    |||||
Db 2224 GCATCGCAGCATT 2210

Query Match      60.0%; Score 15; DB 65; Length 23873;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattt 15
|||||
Db 2224 GCATCGCAGCATT 2210

RESULT 43
AF168792/c
LOCUS
DEFINITION      AF168792 31424 bp DNA VRL 23-SEP-1999
                  Gallid herpesvirus 1 ICp18.5 gene, partial cds; DNA binding
                  protein, DNA polymerase, UL31 homolog, UL32 homolog, UL33 homolog,
                  UL34 homolog, UL35 homolog, very large tegument protein, UL37
                  homolog, capsid assembly protein, ribonucleotide reductase large
                  subunit (VHS), and UL42 homolog genes, complete cds; and UL43
                  homolog gene, partial cds.
ACCESSION      AF168792
VERSION        AF168792.1 GI:5918968
KEYWORDS       gallid herpesvirus 1.
SOURCE         gallid herpesvirus 1.
ORGANISM       Alphaherpesvirinae; Varicelloviruses, no RNA stage; Herpesviridae;
                  Alphaherpesvirinae; Varicelloviruses.
1 (bases, 1 to 31424)
REFERENCE
    Johnson, M.A.
    Sequence of the infectious laryngotracheitis virus (SA-2 strain)
    unique long region UL28 to UL43
    Unpublished
    2 (bases 1 to 31424)
    Johnson, M.A.
    Direct Submission
    Submitted (14-JUL-1999) Division of Animal Health, CSIRO, Private
    Bag 24, Geelong, VIC 3220, Australia
    Location/Qualifiers
    1..31424
    /organism="gallid herpesvirus 1"
    /strain="SA-2"
    /db_xref="taxon:10386"
    complement(<1..752)
    /note="UL28 homolog"
    /codon_start=1
    /product="ICp18.5"
    /protein_id="AAD56200.1"
    /db_xref="GI:5918969"
    /translation="MAEETQLRLMALAGLOLTLLFQIDILKRCDEPILLOKRLKRSRIK
    HNALMALYIHSLLSNELSMQVNYRLTQTCYVWLWLRACSEIAALCNEIDKYSVLK
    RCFEATLGYGKGFICPLHAEVQLQYGMSSIIISKELVINDVFNKQNYCHCV
    SKNAIEALNTIDEFLVSSGSGSLVAAPDYDHGQPCVYCLEELSVTANOGETIYRRL
    GYKIDHLTROPFVNVTDDLKRLHLPFLRDVDP"
    complement(800..805)
    /note="UL29 homolog"
    /codon_start=1
    /product="DNA binding protein"
    /protein_id="AAD56201.1"
    /db_xref="GI:5918970"
    /translation="MEKSSGCTTGGKSDGCPWCIIGVYVYARPREIDGKEWILC
    AKSDQSCATAPLRLGLTVESDFKPNVAAVIGTKSSGVGGNCTAILSPCHFSTTV
    IFHGECIPFTPTNLTKCEDAREHGFSSLPNGPVPNAIETGEETCKSLNMDP
    DKMLYLVAEPFCEAVYVCNTYIHYGVSVELYINSHVTRIPYIPVQMYMPDIALRI
    CRNPFDSNRNIGECIYPTPLFNKALNRILHGATLGTQGLSLTROLEAVARGAAML
    AFDGSEGCGLTSDKTFELAONTQQTGPPKQNVVEERRAACSLAALATRLSVSC
    APYFDNCGCLPYEWPFLCDCKTOMERVOALDFSAELAGIVGAAIFSNNSPLIASE
    VVDGAPPGERASSGLTRFLYLCGLHLLGCPQVDYSGKRWGCTPSPUTPONTPEY
    DPEHLAYACGFSPELLSLIFYLEKSKQLSTRPNMNLISISNGKAPCTGAVPECRM
    CNDSRNHCIRYTMORLTFPIPTMRRSPIAVIGAVDGDYDCTDLGNFAPYSQIK
    RAGDEPAKAVMNDTYRIGWRVQYLVSEGLNKDGTEDHNTISSIAELKTVFEKIQ
    NFVGSECFKSALSGYRHYKHEHLSIVHTFSGSPESYSSCPVLSLCAQTRSLI
    FQDLILSQIHGFTDTPROESKMPRSSAMPALRFAFNGLDKGLSGKYEPFSISVPCV
    NAPUTMRPNTQALAOEYSLSRGOVLKKEFKVKNRVFNGSSTSSCRVLSRPFI
    VMGIMLSKVYGOONSTTVFSGNWSGLICNSGQSVNSALTDNPVKFKALACRRKI
    ILSGGQVASFQENTIASQVRSLLDAGGSPTSIYAVLRVLGECMDITAEETWMTD
    DKYLVNLVLELREDISGQNGSVATAQSMLELEGKDTCPENAGEMLVWADEDEYTN
    TGDISEDIELGAPSAKKPSLSADILFC"
    4010..7033
    /note="UL30 homolog"
    /codon_start=1
    /product="DNA polymerase"
    /protein_id="AAD56202.1"
    /db_xref="GI:5918971"
    /translation="MDTFSSGTAARVAPIIGLPHSGASGYTSVREFPYVCPTCIN
    GGRIRGTQGRATSKPTFTHNEROLDLSTHGAWPVRMKYMNAGPMTHHNKGDYKVF
    YEFVYDLFENTEIAQSCVRMHRSFLDLTKPTGTVITLVGNSACKRVAVHYGSLP
    YFVKRREIDQAAKVTNSBELAHALAKTRKSLKNSPFLVSAESFCIDVVRDRDI
    YSESEDFYRVKVCNGKVMKTCIDNFPSPVKYESNVDAITRELLDNLTSGWYRF
    RAQGGALQIRDPQHATSDVFNCTASNLGNNVSWPDYKLLCFEDICKAGANEF
    AFPSAKVEDLVQISAVYSLTKEKEEITFSGTCSELSDCSDGTVTCACGSEFE
    LLLCFMPTFRQYSPERVSGYNILGDMAYLSNKLKSVYGMRLDYGKANSWGTFKLOD
    PSARGLRFRKKVKNVINGMIDMPTISYEKLKLPVSKLVNAVECVGEORIDLAYKDI
    VMPAAGPKRGKIGEYCLQDSRLSGSLFFPSPHLEMSAVAKACIPLTRALDGGOL
    RYVTCLLQRSTAGSVLPDKKSAFSGSTLASDAPARNVYGAKVLDPEIGFH
    VTPVYVDFASLYPSIIQSNLCLYSTLTHDPAALAGLOPEKDFDIDVQGRFVRE
    HIRKSLSLDLGLDMRKAIRAKITAEETEERILLDKQAALSCLELCYGTGMVH
    GMLPCLVASTVTAIGRMLLTKAHIEKWRSGNQFAEKLFGPGRQLQNIYSVRVI
    NGETDSAYCEVYRSLHNFNSWHLGRIDLNKRFLEVRKLEWFKFAAFLILIGKEK
    YIGVHHGGKMLMKRVLDVLFKTNKFVNTASRLVNLLEDEDEIAAADTARAGFDFD
    CLPSGLTKLRLTAELARLITGNLNIIRDFTIMTAELSRAVDNYSASLKPLHLYQKKA

```

polyA_signal
CDS

APRELPPIKRIEYVITIEPQGVNDPATESFPN"
complement(7202..7207)
complement(7203..8222)
/note="UL31 homolog"
/codon_start=1
/protein_id="AAd56203.1"
/db_xref="GI:5918972"
/translation="MSDFSRRAALIRREGITGAIERHRIYISLYRKRSSRSRANL
DTRRTFFDAFFKLTASPEDTISLIRWTPIVROENISLPPYNVNASFAPGDCITISLE
VYGVPGSCSILCSYGSNAPPLDPALEAFHLHLSVVYFELKSVFSGESV
VSGAYENEFLLDTIKNLAQTALFYAVYVKGSLDSFTVLLVNDPSKRFSGYL
YVFKQSTPLHGLAKQLQKFCNFCGTHWVDVAGFELLIVSKEDPGSEHRRVKL
TPDSYVRRCDLTVDEKIHETKDYTFATISPTTSGMIQMSRCFVQEPQLQSQKI
VFSLF"
complement(8215..9963)
/note="UL32 homolog"
/codon_start=1
/protein_id="AAd56204.1"
/db_xref="GI:5918973"
/translation="MTSKRSTAEETELSKGWKAGVLSKLYGVDPAILFNDKLSDE
LUGAHLIEYHEESTEAQENRIPDIDSFVETMADLYTLDKCAVCKITKYQKF
PVTPEMVDYSLLCFKSOSIPLCATSTFIATPEFIFIMDKHYLHSHSTSLVGAISRRE
FSLDQKXFFNCFKAVEGLDSKIDLNYSFVQSVARYALLSTRYOALRAKIA
PVSASADNFPSGMAVALLKWEYARPLECFCSACMLKRNKTETSAFRTCAQ
KALEDISTESKIVEGSETLTPSKWGFDTLQVKNTRLPKTAQYICDNTKQSLAIG
YDNDNGENFVAPIIRNTVYVNFHATLIQVKNTRLPKTAQYICDNTKQSLAIG
PILASILPHPTSGDGTGCVLNCMLTREHWHALRKLKSVGSCNVNSSLFDGIEP
ALFTFEDYTALDGGRMFLRMAGANAIYKHFCDPLCAVNTLRVNPKVLMWEGQPKD
PELLQYKAEIATANTFOEYRCVGLWILAFKAYQLSPRPPTALNSFIRGAETYLKR
HGISCALAEHALTRYV"
9962..10321
/note="UL33 homolog"
/codon_start=1
/protein_id="AAd56205.1"
/db_xref="GI:5918974"
/translation="MASTVPRRLSELVDEELAEPLSELPRKYIGSAGSPVIEIWFY
LSFPFIEVNAIPTDANKLYLAHTANVAEELRYNLDGKRCMAHAEIARRRREFAQI
LSKPLDHLQILGALPC"
10356..11228
/note="UL34 homolog"
/codon_start=1
/protein_id="AAd56206.1"
/db_xref="GI:5918975"
/translation="MVLRIQKIIKAMPQCARNGDKYCOLVRFVNAGLSLAGTTA
SLVYTRNARLAPDGIPTLLAKLDGPPIPAETYLEAMNSFLNIGESWIRIQTQAV
IVAGCFTKTNCKDQIPASSPTISLAAKSLWSSSVKEMKQKLOKLNAPLATMYE
ISFVGRSDTITIRFAYRSDSEPNLIKIRRVQEAIDSAEQEDSYKTAALLKKTME
TTSEMTSHSVAESKPSRSIFERVYTRIVHVKCLNPRFSLQPLMLVLMGVAVPAFL
TLIFYLIRVQATS"
11254..11628
/note="UL35 homolog"
/codon_start=1
/protein_id="AAd56207.1"
/db_xref="GI:5918976"
/translation="MSQSNHSTPQDVRVNAEAEMLKAFLLTPDRPADPYMOTLN
GITLDQLEVARALKNEIPREDTVRONLRNNTTLALMALNAPMLRDLNLRPFAR
LSGPQGLWTFVGRQRTPRPRR"
complement(11653..19323)
/note="UL36 homolog"
/codon_start=1
/product="very large tegument protein"
/protein_id="AAd56208.1"
/db_xref="GI:5918977"
/translation="MNLDDTICVWRPPPCGGANATGSDMSLADSLLDILDOGAIE
TRTGKAVEQDKSKCSFMDICHELPLRMSSWDEPDLMLIPSIPYGNVGYDMEVIE
GTPSCNMIRGTTKIAKIRKGLFTIMIGRGVAAVCGGKDDPVYIFDSQWEMSS
AYVQLPRIDLPEFVAYVKNMGSVAIDMNFYWNADGWDEKVEKTLQEIITAA
VMASTGASDAELSPVLTSLQIIFPHFAYALGNLEFESKAKNLELADRRKAIAPWS
PVRPSKKRQPKPNKDQIGSSEKLTNLPKROKHSVNSDPEAGSNTPKPSARKL
WNLTITADTNPSATIEIGSTPIKNALEDPLKLPAPQEPFPAFLRSLRILREK
VNTCNRLFFIENGYSKTNLESTVDLLDPLATLLARIGEKNIENIITTLRLRYQD
LIEKKHLLLECHNSGGSVLLRKTOQVSLTOTOEILIAKINOIADSARDQLPT
AYEQLSELLECSNKTSLIEDVNSHGKLTDDVVERAAKSVIESEAFADAKTAVLSL

polyA_signal
CDS

LANWTVIDETICNVGINQATQATPDIPQAIIQGLRKFTGLDQVQDINRKLDDAIIR
IEDAAVKFKKATAGAIIDISISALSSGGVGLGRESIPTLFLVENAVNWKNIIEERRE
QLIRQAGTSPKNIOVLEFRKPLRIFDLVLDGKDLTDEKLRKSLHLFEAREAGSLY
PSPTKIWIHIDINAQSGFLTENMLAELENEVEATAPPOVLRVLTQVHLHQBESI
KVAKAARTTVEQSRRLDAEDRSLIDQFSKEIKETAEHTKRTVTQILRLPKKF
EGLRNIKSALOTNRNPOBEVINLARNDDFIGKRVDDFENLEETKRTVQTSGLANTDLH
LDVDAIETATILLSGKNKHALNRAFEVLDHAERLLFVIETAKANGADGDFGRAROE
VEAVLVONNRGASQIPQFLESIRDNLYELEYSEAYORARKKOLSDVIERLTDILNA
TEKLVNPELRQIADSAIVILVERVKEVAOLDITNEERSPELLSLEKACSNISEAA
HNAIIAIEIEMLEATKAALQDGRSACFSADDWETIIDLPHVNSAHKHAHTELKHLA
QNVVNDLATETKLLRSPFSPNYSAGNAOGKNKIAVAPIAASVTWHSPLVSKYKYE
TLFGAKVEGLTIVCAIKADILETAETVVEGQIDTVGVFRLSQMLRIPELSEFVDF
YLOAHEOFTESSKIOTIKERAILLODELKELQALAEASQIERSPENAKRRLDHGEV
IPTALIAKELPDLSDDDTVDEHTAYQDASHRALRDVLFLSAETINAEKPOOKNTD
GKIKNMLQLSWREEVLRMQGLKLLGLLGRVKEVAPGIERAESLEDDIILFL
NLFREVEKSNSTGITSVDPKLEWMPSSAHLDTCELVNSIDERGPLEDFKKRLDEL
KRYEEADAGKDLKLEAYYKFLSAATSEKTSKKEKDTSEESGAYNAVHNLKESRE
RLASRLQVQAVTSTPEIKKYQKDISETPILOKAEENIITLMNARTFDEVLFKAK
TMHFKALKLSHLELTAFESVPKWRKAYEDIKKILIRLELYLAYKALGIAFSDDI
IPRLSAAETPPSSRRELARSSYRONKKAIVMTVPHLCTIGSVWVTDNGMPLEYT
LCYRTATAKAAALNSAALCTSSDOSEPLSLEENPKNELVKLSVRONAALSVAAT
LWMAKTSIDIEAGFSYITFCSTRTWPSNQTDTPMVAACITYSALTYLTLSKTLGS
ARDVIIDENGNFIPTDLDKINQKQDFKAVIYITAIIDILILLITACEPAHLVYFOR
LNYLQTEYIVNTLDVLSRAVREKVGITSLDPYQRTONKKAISFPLSKGRQHFDS
AGACFVIDPKMLSQPPSSILRFLDPKMRFPVSRSGICMEKFCSEGETSAFLMI
LALKCIPRILEALWVSMGPHDDIDGEPIDTMHYVWSRSPSCDYDFSPNDSLRVF
EAPDKLGTVDGRAGIYSLREEPKHSIMDLTAFDIAITSLFGAPVVIYABAGLSSN
NGLVLFSTIFTDGRKNPLLESASSTATAIQLDIIAGDANPTENACLASQIAELST
VSSKPLKAPSLFAIDRGNKITOILVSELTAPSPFVIRLHGDPKYEDLVPKFSDTG
TSSNGLDGNPDYTSYVSAVPSQSGTESSEFEQVLETCMDKIQSPSITSDSNDF
AVITDETLEHPTDPEKDRSHLARTKTSREHGPPESTSHRDKHRTAEAKQTHCS
NEDIAHEPTYSVTWKDEPPGESDEDETPAYYVKGSHKSHREESRRADHKKGRIE
EGELVYRDVQALDLYLRKLNDALRQCTITPYLKSITQTPDSSSEDDAVNEISSGTSI
DKRAPSIEDAESAESKACERELSPLPRTASDIYASLSVTCRMVLKIGIKHARDYAL
DQQAIRADETHKIKIILG"
complement(19460..19465)
complement(19948..22620)
/note="UL37 homolog"

polyA_signal
CDS

Query Match 60.0%; Score 15; DB 58; Length 31424;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgactct 22
|||||
DB 13467 CAGCATTTCGATCT 13453
|||||

RESULT 44
CEY62E10A 57237 bp DNA INV 09-NOV-2000
LOCUS Caenorhabditis elegans cosmid Y62E10A, complete sequence.
DEFINITION Caenorhabditis elegans
ACCESSION AL132865.1 GI:5801580
VERSION AL132865.1
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 57237)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
JOURNAL 99069613
MEDLINE The C.elegans Sequencing Consortium.
REMARK Erratum: [[published errata appear in Science 1999 Jan
1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
3;285(5433):1493]]
REFERENCE 2 (bases 1 to 57237)
AUTHORS Lloyd,C.R.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1999) Nematode Sequencing Project, Sanger Centre,


```

FCSDKPAAVNWIEGRKSVVCEAVIRGEIVNKVLTQVAALVELNMLKLNAGSAVAG
SLGGFNHASNIVSAVFIATGODPAQNVSSOCITMWEAINDGKDIHISTYMPISIEVG
TVGGTOLASQACNLGLGVKGASTESPGRNARRLATIAGAVLAGELSLSMAIAAQQ
LVRSHMYKNSRRDISGATTTTTTTT
complement(7258..7283)
/rpt_family="(CAA)n"
complement(11154..11209)
/rpt_family="AT-rich"
complement(11551..11580)
/rpt_family="AT-rich"
complement(11933..12004)
/rpt_family="AT-rich"
join(<12390..12391,12491..12629,12730..12861,12929..13019,
13099..>13268)
/gene="F15M4.2"
12390..13268
/gene="F15M4.2"
/note="similar to unknown protein GB:AAD30603 [Arabidopsis
thaliana]"
join(12390..12391,12491..12629,12730..12861,12929..13019,
13099..13268)
/gene="F15M4.2"
/codon_start=1
/product="unknown protein; 12390-13268"
/protein_id="AAF16653.1"
/db_xref="GI:6554471"
/translation="MFDVKALXVKGELRETLPQSGRVYKQLGKNSWYEVKISY
PASIPARVQLQKHEMELKLNMRLLNTEKLIFKAESLREVNNKAGLNVLVLEP
EGIVAPNSRERVFIYINIVCEBGLMGIPYSSVLLVALCLVSVFVPRSLPSLL
TKDQGLRSHRSGKDS"
join(<14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..>15506)
/gene="F15M4.3"
14056..15506
/gene="F15M4.3"
/note="similar to cinnamoyl-CoA reductase GB:CAA56103
[Eucalyptus gunnii]"
join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)
/gene="F15M4.3"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 14056-15506"
/protein_id="AAF16654.1"
/db_xref="GI:6554472"
/translation="MAYKQKVCVTGAGGFASLWLVKFLLSRGYTVHGTVRDPCDEKND
HUKLDNASKNLKFKADLDDDEGLFSAIDCGVFHIAFPVEGEELIKPALTGK
NVLEACTETKVQVVSIAAVYNPKVPQDVAKDCWSDTQVHLSELEGWRYVL
AKTLTERALEWSKRNFAVWVTLCPSVIIGPRLSTLNSSSLGLKFKGKISLSD
ELYLDVRDVALDALLVYENREATGRYICNHSLSYTDLSMEKLNKMYPRRNPESFTE
VKEKEVRPLSAEKLKNGKFRPLEETIDDSVVSFEAAGDLFKA"
16281..16361
/rpt_family="(CAT)n"
join(<16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..>18423)
/gene="F15M4.4"
16955..18423
/gene="F15M4.4"
/note="contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
join(16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..18423)
/gene="F15M4.4"
/codon_start=1
/product="putative RNA-binding protein; 16955-18423"
/protein_id="AAF16655.1"
/db_xref="GI:6554473"
/translation="MAYQVPVPGSGFHYLNSPFGDTTFTKVFVGLAWETOSETLRQHF
EQYGEILEAVVIADKNTGRSKYGFVTFROPEARRACADPTPIIDRRANCNLASIG
RPRPLPYAVIPNMDLLPRTLEMCVLEVHLEIIRISNHFHITTSKELSLIMVTAY
GPEYWSQSGSLSPYMGQQLQVYVPGAVNSPVYQGLSQITIPNGHYTAVQGYIS
VPGSHILQLGGPTVSTMTSSMPALQAPYSPGPGAPVQSHIIVHSPQMQSTASDQ
TTR"
complement(join(<19352..19378,19455..19499,19581..19699,

```

```

19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))
/gene="F15M4.5"
complement(19352..20843)
/gene="F15M4.5"
complement(join(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))
/gene="F15M4.5"
/codon_start=1
/product="unknown protein; 20843-19352"
/protein_id="AAF16656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSSPQSFNPRVTITRRRCSSIAAISVLDSSNEEQ
HRISSRDHVGMRDVMLOIASVVFLLPALSPAFETNASEAFRVYTDENKFEISI
PQVSIATITGLGPDFTRMESFGKVEAFETLVSLGDRSMQKPVGTAKLIDRSKSGFY
YIEVTLQNPGEARKHLYSAIGMATNGWYNRLYTVTGQFTDESAEQSSKIQKTVKSR
FI"
complement(20998..21026)
/rpt_family="AT-rich"
complement(21140..21167)
/rpt_family="AT-rich"
join(<22462..22639,22837..>23090)
/gene="F15M4.6"

Query Match          60.0%; Score 15; DB 12; Length 94487;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

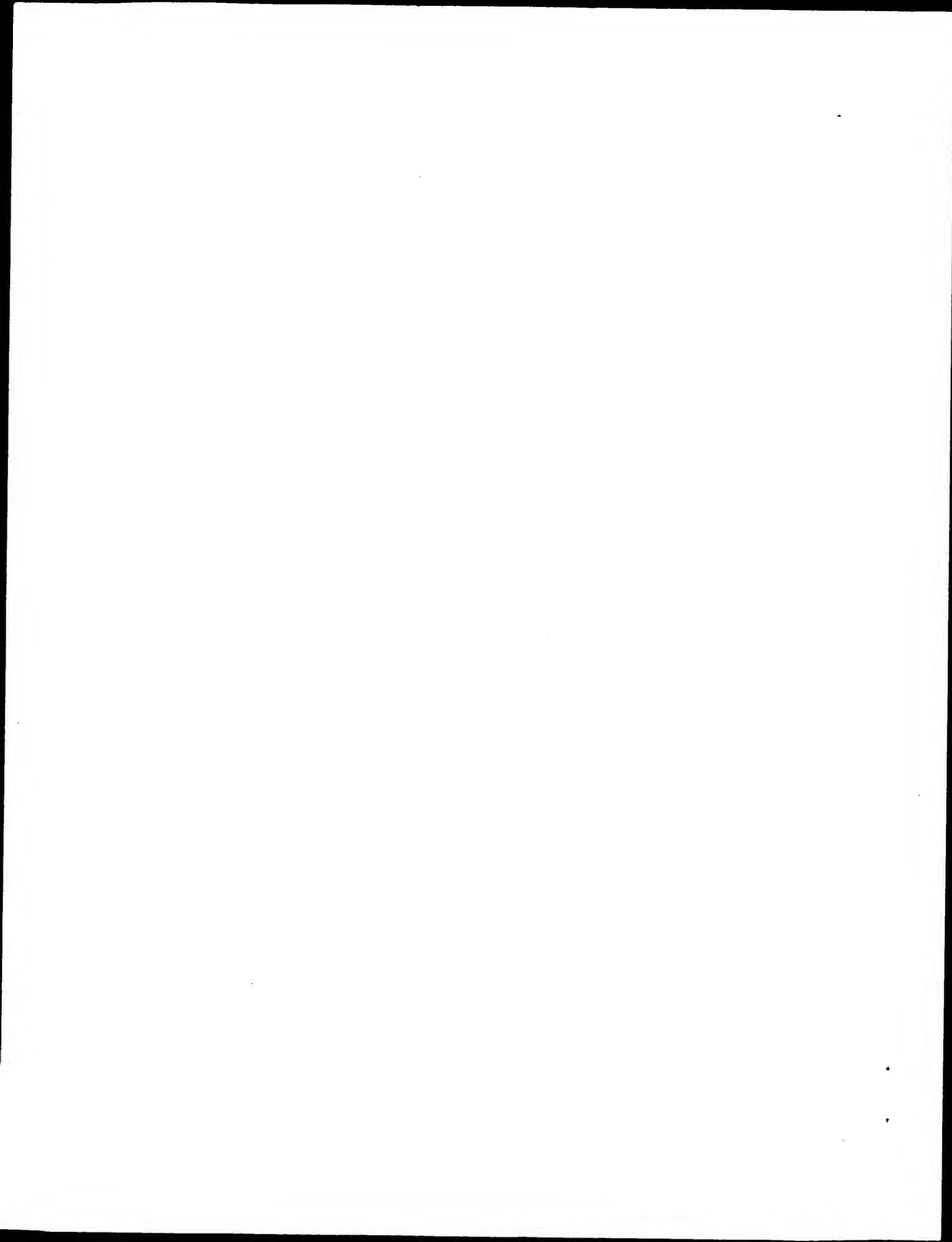
Qy 4 tcgccagcatttcga 18
      |||||
Db 17074 TCGCCAGCATITCGA 17088

Search completed: October 9, 2001, 15:47:19
Job time: 14825 sec

```


us-09-396-196f-7.oli.rge

Wed Oct 10 07:46:02 2001



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:31 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 24498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E.coli Bio B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAO62386	Biotin-biosynthesi
5	15	60.0	5773	21 AAR76014	Human ORFX ORF1569
6	14	56.0	487	21 AAC74884	Human ORFX ORF439
c 7	14	56.0	549	21 AAC93854	Cat flea hindgut a
c 8	14	56.0	1344	21 AAZ56381	Escherichia coli f
c 9	14	56.0	4189	21 AAZ49334	Murine multidrug r
10	14	56.0	4313	14 AAO38950	Mouse multidrug re
c 11	14	56.0	5319	19 AAV35235	R. prowazekii S-la

c 12	14	56.0	143068	21 AAF21105	Human low adenosin
c 13	14	56.0	143068	21 AAF21272	Human low adenosin
c 14	14	56.0	143068	21 AAA34983	Human adenosine re
c 15	14	56.0	143068	21 AAA35150	Human adenosine re
c 16	14	56.0	149412	21 AAA35151	Human adenosine re
c 17	14	56.0	152740	21 AAF21273	Human low adenosin
c 18	13	52.0	22	19 AAV11150	Oligonucleotide #3
c 19	13	52.0	51	21 AAA76686	Human clone cg2838
c 20	13	52.0	51	21 AAA76687	Human clone cg2838
c 21	13	52.0	61	20 AAZ19606	Complement system
c 22	13	52.0	300	20 AAZ14744	Human gene express
c 23	13	52.0	322	21 AAT27962	Plant microsatelli
c 24	13	52.0	447	17 AAT27962	Hepatitis C virus
c 25	13	52.0	591	21 AAC54077	Arabidopsis thalia
c 26	13	52.0	606	21 AAC35502	Arabidopsis thalia
c 27	13	52.0	629	21 AAF09259	Fusarium venenatum
c 28	13	52.0	714	20 AAF17005	Human gene express
c 29	13	52.0	725	22 AAF22436	Human breast cance
c 30	13	52.0	750	20 AAZ17191	Human gene express
c 31	13	52.0	974	21 AAC35170	Arabidopsis thalia
c 32	13	52.0	974	21 AAC48947	Arabidopsis thalia
c 33	13	52.0	985	21 AAC46920	Arabidopsis thalia
c 34	13	52.0	996	21 AAC47663	Arabidopsis thalia
c 35	13	52.0	1007	19 AAZ96443	S. pneumoniae deri
c 36	13	52.0	1038	21 AAF12052	Aspergillus oryzae
c 37	13	52.0	1065	20 AAX84999	Human secreted pro
c 38	13	52.0	1110	18 AAT84197	DNA encoding a UDP
c 39	13	52.0	1114	21 AAC41953	Arabidopsis thalia
c 40	13	52.0	1179	21 AAL15961	Human protein clon
c 41	13	52.0	1226	21 AAC36239	Arabidopsis thalia
c 42	13	52.0	1263	9 AAN82205	Nad B gene encodin
c 43	13	52.0	1287	21 AAG93673	Human peroxisome a
c 44	13	52.0	1291	21 AAG93674	Human peroxisome a
c 45	13	52.0	1313	21 AAC61879	CDNA encoding a hu

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
XX 12-APR-1999 (first entry)
DT
DE E. coli biotin synthetase (BioB) coding sequence.
XX
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
KW
XX Escherichia coli.
OS
XX
XX US5869719-A.
XX
XX 09-FEB-1999.
PD
XX
XX 30-APR-1997; 97US-0846338.
PF
XX
XX 30-APR-1997; 97US-0846338.
PR
XX 08-MAR-1995; 95US-0401068.
PR
XX (NOVS) NOVARTIS FINANCE CORP.
FA
XX Patton DA;
PI
XX WPI; 1999-152902/13.
DR
XX P-PSDB; AAW73906.
DR
XX Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```

XX PS Example 2; Column 37-40; 34pp; English.
XX CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
XX CC be used in the transgenic plant of the invention. The transgenic plant,
XX CC plant cell or plant tissue is transformed with a chimeric gene encoding
XX CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
XX CC produces more biotin than a non-transgenic plant, cell or tissue. The
XX CC plant is used as an improved dietary source of biotin (vitamin H) for
XX CC humans or animals.
XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match      100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcgcagcatttcgactcctg 25
   |||||
Db 90 gcacgcgcagcatttcgactcctg 114

RESULT 2
AAN91329
ID AAN91329 standard; DNA; 1084 BP.
AC AAN91329;
XX
XX
XX 15-FEB-1990 (first entry)
DT
DE E.coli Bio B gene.
KW E.coli; Bio B gene; biotin.
XX Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 24..1064
   /*tag=a
XX
XX GB2216530-A.
XX
XX 11-OCT-1989.
XX
XX 17-MAR-1989; 89GB-0006210.
XX
XX 22-MAR-1988; 88GB-0006804.
XX 17-MAR-1989; 89GB-0006210.
XX
XX (UKAG-) UK MIN. AGRIC. FISH.
XX
XX Pearson BM, McKee RA;
XX
XX WPI; 1989-295085/41. P-PSDB P91392
XX
XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
XX - derived from E.coli and capable of replication and expression in other
XX microorganisms, esp. yeast.
XX
XX Table 3; page 33-4; 52pp; English.
XX
XX The gene can be used in a plasmid for expression of enzymes of the biotin
XX synthetic pathway. Pref. control sequences for expression in S.cerevisiae
XX are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
XX Lactobacillus. Insertion of bio B improves biotin yields in
XX microorganisms which export biotin, or enables growth in media contg.
XX little or no biotin of organisms unable to synthesise biotin for their
XX own use.
XX
XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match      100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcgcagcatttcgactcctg 25
   |||||
Db 113 gcacgcgcagcatttcgactcctg 137

RESULT 3
AAN60496
ID AAN60496 standard; DNA; 1121 BP.
AC AAN60496;
XX
XX 17-OCT-1991 (first entry)
DT
DE Sequence encoding biotin synthesising enzyme.
XX Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH Key Location/Qualifiers
FT CDS 42..1082
   /*tag=a
XX
XX JP61149091-A.
XX
XX 07-JUL-1986.
XX
XX 24-DEC-1984; 84JP-0272605.
XX
XX 24-DEC-1984; 84JP-0272605.
XX
XX (NIPS ) NIPPON SODA KK.
XX
XX WPI; 1986-216622/33.
XX P-PSDB; AAP60536.
XX
XX Double stranded DNA encoding biotin synthesising enzyme -
XX comprises transformed mutant E.coli strain contg. cyclic doubled
XX stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
XX Disclosure; Page 534; 23pp; Japanese.
XX
XX The sequence may be expressed by a transformed E.coli host, cultured
XX in a medium containing desthiobiotin.
XX
XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match      100.0%; Score 25; DB 7; Length 1121;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcgcagcatttcgactcctg 25
   |||||
Db 131 gcacgcgcagcatttcgactcctg 155

RESULT 4
AAQ62386
ID AAQ62386 standard; DNA; 5872 BP.
AC AAQ62386;
XX
XX 16-NOV-1994 (first entry)
DT
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX Biotin; expression; enterobacteria; vitamin H; synthesis;
XX plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
XX promoter ptac; biotin synthase; KAPA synthase;
XX 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
XX

```

```

KW dethiobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX seborrhea; dermatitis; ds.
OS Escherichia coli DSM498.
XX
FH Key Location/Qualifiers
FT promoter 1..96
FT /*tag= a "promoter ptac"
FT /evidence= EXPERIMENTAL
FT 23..28
FT /*tag= b
FT /standard_name= "promoter ptac"
FT 45..50
FT /*tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter ptac"
FT 105..109
FT /*tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "bioB RBS no. 9"
FT 117..1157
FT /*tag= e "biotin synthase"
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT 1141..1146
FT /*tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /*tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /*tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /*tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT 3030..3033
FT /*tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /*tag= k
FT /EC_number= 6.3.3.3
FT /product= "DPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "dethiobiotin synthase"
FT 3712..3750
FT /*tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
FT /*tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /*tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FT RBS 5088..5093
FT /*tag= o
FT /standard_name= "ORFI RBS"
FT 5098..5574
FT /*tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORFI"
FT /number= 6
FT 5583..5644
FT /*tag= q
FT /standard_name= "rho-independent transcriptional
FT terminator"
FT 5583..5605
FT /*tag= r
FT
FT W09408023-A.
FT
FT 14-APR-1994.
FT
FT 01-OCT-1993; 93WO-EP02688.
FT
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT
FT (LONZ ) LONZA AG.
FT
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT
FT WPI: 1994-135587/16.
FT P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT
FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT bioB, bioF, bioC, biob and bioA genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,
FT loss of appetite and tiredness.
FT
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT
FT Query Match 100.0%; Score 25; DB 15; Length 5872;
FT Best Local Similarity 100.0%; Pred. NO. 2.1e-05;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 1 gcacgcgcagcatttcgactcgtcgt 25
FT |||||
FT Db 206 gcacgcgcagcatttcgactcgtcgt 230
FT
FT RESULT 5
FT AAC76014/C
FT ID AAC76014 standard; cDNA; 5773 BP.
FT
FT AAC76014;
FT
FT 08-FEB-2001 (first entry)
FT
FT Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
FT
FT Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
FT vulneray; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
FT anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
FT immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
FT hypotensive; dermatological; immunosuppressive; antiinflammatory;

```

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB41805.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2354-2357; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
 Query Match 60.0%; Score 15; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 catgccagcatttc 16
 |||||
 Db 4886 CATGCCAGCATTTC 4872
 RESULT 6
 AAC74884
 ID AAC74884 standard; cDNA; 487 BP.

XX AAC74884;
 AC
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF439 polynucleotide sequence SEQ ID NO:877.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB40675.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 899-900; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 487 BP; 122 A; 99 C; 129 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 487;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattgatcc 21
Db 316 cagcattgatcc 329
|||||

RESULT 7
AAC93854/C
ID AAC93854 standard; cDNA; 549 BP.

XX AC AAC93854;

XX DT 19-FEB-2001 (first entry)

XX DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:349.

XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.

XX OS Ctenocephalides felis.

XX PN WO200061621-A2.

XX PD 19-OCT-2000.

XX PF 07-APR-2000; 2000WO-US09437.

XX PR 09-APR-1999; 99US-0128704.

XX PA (HESK-) HESKA CORP.

XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;

XX WPI; 2000-656323/63.

XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -

XX PS Claim 26; Page 361; 964pp; English.

XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention.

XX SQ Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 549;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattt 15
Db 421 CATGCCAGCATT 408
|||||

RESULT 8

AAZ56381/C
ID AAZ56381 standard; DNA; 1344 BP.

XX AC AAZ56381;

XX DT 17-MAR-2000 (first entry)

XX DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX KW Flagellin; fliC; antigen; detection; ds.

XX OS Escherichia coli.

XX PN WO9961458-A1.

XX PD 02-DEC-1999.

XX PF 21-MAY-1999; 99WO-AU00385.

XX PR 21-MAY-1998; 98AU-0003634.

XX PA (UNSY) UNIV SYDNEY.

XX PI Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc. -

XX PS Claim 3; Page 225; 245pp; English.

XX CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
CC represent primers used in the exemplification of the present invention.

XX SQ Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;

Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatt 14

Sequence 4189 BP; 1204 A; 875 C; 1028 G; 1082 T; 0 other;

Sequence 142068 BP: 41194 A: 30122 C: 32403 G: 39349 T: 0 other:

Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgattcc 21
|||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 13
AAF21272/c
ID AAF21272 standard; DNA: 143068 BP.

AC AAF21272;
XX
XX
DT 14-MAR-2001 (first entry)
XX
XX
DE Human low adenosine antisense oligonucleotide related sequence #2839.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
XX WPI: 2000-679539/66.
XX
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
XX Disclosure: Page 1186-1219; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors and immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or

CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
XX Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgattcc 21
|||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 14
AAA34983/c
ID AAA34983 standard; DNA: 143068 BP.

XX AAA34983;
XX

DT 28-JUL-2000 (first entry)
XX

DE Human adenosine receptor related polynucleotide SEQ ID NO:2672.
XX

KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.

PN WO200009525-A2.
XX

PD 24-FEB-2000.
XX

PF 03-AUG-1999; 99WO-US17712.
XX

PR 03-AUG-1998; 98US-0095212.
XX

PA (UYEC-) UNIV EAST CAROLINA.
XX

PI Nyce JW;
XX

DR WPI: 2000-205971/18.
XX

XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX

PS Disclosure: Page 851-882; 1343pp; English.
XX

CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic

CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONS reduces side effects. The A-containing ONS break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX SQ Sequence 143068 BP; 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcatttcgattcc 21
 |||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 15
 AAA35150/c
 ID AAA35150 standard; DNA; 143068 BP.

XX AC AAA35150;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers

XX PS Disclosure; Page 1106-1138; 1343pp; English.

XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or

CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytotatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONS reduces side effects. The A-containing ONS break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX SQ Sequence 143068 BP; 41194 A; 30126 C; 32402 G; 39346 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcatttcgattcc 21
 |||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 16
 AAA35151/c
 ID AAA35151 standard; DNA; 149412 BP.

XX AC AAA35151;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:25.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or

PT cancers -
XX Disclosure; Page 1138-1171; 1343pp; English.
XX

CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC the breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX

XX Sequence 149412 BP; 43049 A; 31388 C; 33852 G; 41123 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 21; Length 149412;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcatttcgatcc 21
|||||
Db 96173 CAGCAATTCGATCC 96160

RESULT 17
AAF21273/c
XX
XX AAF21273;
XX
XX 14-MAR-2001 (first entry)
XX
XX Human low adenosine antisense oligonucleotide related sequence #2840.
XX
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX human; airway disorder; bronchoconstriction; lung inflammation;
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
XX respiratory obstruction; pulmonary obstruction; impeded respiration;
XX surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX cancer; ss.
XX
XX Homo sapiens.
XX
XX WO2000062736-A2.
XX
XX 26-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX

PA (UYEC-) UNIV EAST CAROLINA.
PA PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
XX WPI; 2000-679539/66.
XX

Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
XX Disclosure; Page 1219-1254; 1592pp; English.
XX

CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with the
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors and
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX

XX Sequence 152740 BP; 44169 A; 32023 C; 34549 G; 41999 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 21; Length 152740;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcatttcgatcc 21
|||||
Db 96173 CAGCAATTCGATCC 96160

RESULT 18
AAV11150/c
XX
XX ID AAV11150 standard; DNA; 22 BP.
XX
XX AAV11150;
XX
XX 14-JUL-1998 (first entry)
XX
XX Oligonucleotide #3 for nucleic acid assay.
XX
XX Nucleic acid assay; hybridising; probe; target; duplex; detection; ss.
XX
XX Synthetic.
XX
XX DE19633436-A1.
XX
XX 26-FEB-1998.
XX

CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
CC consecutive pairs of nucleotides containing SNPs which result in changes
CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
CC in non-conservative changes. The SNPs in sequences 1187 to 1192
CC (AAA77504-A77509) generate frameshift mutations. The invention also
CC relates to a method of detecting a polymorphic site in a nucleic acid and
CC a method of determining the relatedness of two nucleic acids. It also
CC encompasses peptides containing polymorphic sites, antibodies raised
CC against such peptides, and a method of detecting polymorphic
CC proteins/peptides using the antibodies. The nucleic acids are useful for
CC gene therapy of an individual having, suspected of having, or at risk of
CC developing a pathological condition due to the presence of a sequence
CC polymorphism. Such treatment would comprise administration of the
CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
CC peptides can also be used in the treatment of such individuals.
XX
SQ Sequence 51 BP; 13 A; 11 C; 15 G; 12 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cgccagcattcg 17
| | | | | | | | | |
Db 11 cgccagcattcg 23

RESULT 20
AAA76687
ID AAA76687 standard; cDNA; 51 BP.
XX
AC AAA76687;
XX
DT 16-NOV-2000 (first entry)
XX
DE Human clone cg28389525 polymorphic site, SEQ ID NO:370.
XX
KW Human; single nucleotide polymorphism; SNP;
KW detection; identification; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (26,T)
FT /*tag= a
XX
PN WO200029623-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-US27293.
XX
PR 17-NOV-1998; 98US-0109024.
PR 16-NOV-1999; 99US-0109024.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2000-387826/33.
XX
PT Human nucleic acids containing single nucleotide polymorphisms, useful
PT for treating a subject suffering, or at risk from a pathology due to
PT the presence of a sequence polymorphism -
XX
PS Claim 1; Page 270; 543pp; English.
XX

PF 20-AUG-1996; 96DE-1033436.
XX
PR 20-AUG-1996; 96DE-1033436.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Bergmann F, Herrmann R, Kobold U;
XX
DR WPI; 1998-146248/14.
XX
XX Nucleic acid assay - based on mass of probe-target duplex
XX
XX Disclosure; Fig 1; 9pp; German.
XX
CC AAV11148-V11151 are oligonucleotides used in a novel nucleic acid assay.
CC This assay involves hybridising a probe to a target sequence of the
CC nucleic acid to form a duplex and degrading the nucleic acid to form
CC a fragment of a defined length containing the duplex, which is detected
CC on the basis of its mass. This method enables different nucleic acids to
CC be detected simultaneously does not require the use of labelled probes.
XX
SQ Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;

Query Match 52.0%; Score 13; DB 19; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatt 14
| | | | | | | | | |
Db 18 CATCGCCAGCATT 6

RESULT 19
AAA76686
ID AAA76686 standard; cDNA; 51 BP.
XX
AC AAA76686;
XX
DT 16-NOV-2000 (first entry)
XX
DE Human clone cg28389525 polymorphic site, SEQ ID NO:369.
XX
KW Human; single nucleotide polymorphism; SNP;
KW detection; identification; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (26,C)
FT /*tag= a
XX
PN WO200029623-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-US27293.
XX
PR 17-NOV-1998; 98US-0109024.
PR 16-NOV-1999; 99US-0109024.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2000-387826/33.
XX
PT Human nucleic acids containing single nucleotide polymorphisms, useful
PT for treating a subject suffering, or at risk from a pathology due to
PT the presence of a sequence polymorphism -
XX
PS Claim 1; Page 269; 543pp; English.
XX

Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences which contain single nucleotide polymorphisms (SNPs). Sequences 1 to 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are consecutive pairs of nucleotides containing SNPs which result in changes in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result in non-conservative changes. The SNPs in sequences 1187 to 1192 (AAA77504-A77509) generate frameshift mutations. The invention also relates to a method of detecting a polymorphic site in a nucleic acid and a method of determining the relatedness of two nucleic acids. It also encompasses peptides containing polymorphic sites, antibodies raised against such peptides, and a method of detecting polymorphic proteins/peptides using the antibodies. The nucleic acids are useful for gene therapy of an individual having, suspected of having, or at risk of developing a pathological condition due to the presence of a sequence polymorphism. Such treatment would comprise administration of the wild-type nucleic acid sequence. Antibodies raised against polymorphic peptides can also be used in the treatment of such individuals.

XX Sequence 51 BP; 13 A; 12 C; 15 G; 11 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cgcgcagcttcg 17
|||||

Db 11 cgcgcagcttcg 23

RESULT 21

AAZ19606
ID AAZ19606 standard; RNA; 61 BP.

AC AAZ19606;

DT 08-NOV-1999 (first entry)

DE Complement system protein C5 RNA binding ligand 13.

XX Complement system protein; Clq; C5; C3; RNA ligand; neurotropic; stroke;
KW neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW diagnostic; prevention; treatment; complement protein-related disease;
KW Alzheimer's disease; renal disease; transplant rejection; asthma; ss.

OS Synthetic.

XX WO9941271-A1.

XX 19-AUG-1999.

PF 05-FEB-1999; 99WO-US02597.

PR 29-SEP-1998; 98US-0163025.

PR 12-FEB-1998; 98US-0023228.

PA (NEXS-) NEXSTAR PHARM INC.

XX Biesecker G, Gold L;

XX WPI; 1999-527357/44.

PT New Nucleic Acid Ligand to complement protein C5, useful in the
PT treatment of Alzheimer's disease, renal diseases, transplant
PT rejection, stroke and asthma

PS Example 4; Page 42; 120pp; English.

XX This invention describes novel purified and isolated non-naturally
CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.

CC The products of the invention have neurotropic, neuroprotective,
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
CC new ligands are useful as diagnostic agents, and pharmaceutical agents
CC for prevention and treatment of complement protein-related diseases,
CC including Alzheimer's disease, renal diseases, transplant rejection,
CC stroke and asthma.

XX Sequence 61 BP; 11 A; 18 C; 21 G; 11 U; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 61;
Best Local Similarity 69.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgac 20

Db 23 cagcauuucgac 35

RESULT 22

AAZ14744/C
ID AAZ14744 standard; cDNA; 300 BP.

AC AAZ14744;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:2213.

XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX WO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080566.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
XX differentially expressed in different cell types

XX Claim 1; Page 1159; 2479pp; English.

XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and

CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX SQ Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
 |||||
 Db 174 GCATGCCAGCAT 162

RESULT 23.
 AAA31174/c

ID AAA31174 standard; DNA; 322 BP.

XX AC AAA31174;

XX DT 05-JUL-2000 (first entry)

XX DE Plant microsatellite marker #135.

XX KW Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.

XX OS Eucalyptus grandis.

XX PN W09967421-A1.

XX PD 29-DEC-1999.

XX PF 25-JUN-1999; 99WO-N200092.

XX PR 25-JUN-1998; 98US-0105307.

XX PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Havukkala IJ, Bloksberg LN, Glenn M;

XX DR WPI; 2000-116958/10.

XX PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -

XX PS Claim 1; Page 116; 392pp; English.

XX CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

SQ Sequence 322 BP; 43 A; 101 C; 88 G; 89 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atgccagcat 15
 |||||
 Db 21 ATGCCAGCAT 9

RESULT 24

AAT27962

ID AAT27962 standard; DNA; 447 BP.

XX AC AAT27962;

XX DT 11-MAR-1997 (first entry)

XX DE Hepatitis C virus type 10a isolate NN98 bases 478-925.

XX KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection; ss.

XX OS Hepatitis C virus.

XX PN W09613590-A2.

XX PD 09-MAY-1996.

XX PF 23-OCT-1995; 95WO-EP04155.

XX PR 28-JUN-1995; 95EP-0870076.

XX PR 21-OCT-1994; 94EP-0870166.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Stuyver L;

XX DR WPI; 1996-251460/25.

XX DR P-PSDB; AAR96551.

XX PT Hepatitis C virus polynucleic acid unique to unidentified subtype
 PT - used to develop probes and primers for new subtypes and vaccines
 PT to prevent and treat infection

XX PS Claim 6; Fig 3; 150pp; English.

XX CC The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
 CC genome. This sequence represents nucleotides 478-925 from the HCV type
 CC 10a isolate NF98.

XX CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.

XX SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 3 other;

Query Match 52.0%; Score 13; DB 17; Length 447;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatt 14
 Db 406 catgccagcatt 418
 RESULT 25
 AAC54077
 ID AAC54077 standard; DNA; 591 BP.
 AC AAC54077;
 XX 18-OCT-2000 (first entry)
 DT Arabidopsis thaliana DNA fragment SEQ ID NO: 76608.
 DE Arabidopsis thaliana.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 07-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148584.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149829.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157553.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 591;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 attcgatctctcg 24

Db 144 attcgatctctcg 156

RESULT 26

AAC35502
 ID AAC35502 standard; DNA; 606 BP.
 XX
 AC AAC35502;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10429.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 attcgatctctcg 24
Db 159 attcgatctctcg 171

RESULT 27
AAAF09259
ID AAFA09259 standard; cDNA; 629 BP.
XX
AC AAFA09259;
XX
DT 13-MAR-2001 (first entry)
XX

```

DE XX Fusarium venenatum EST SEQ ID NO:1782.
KW XX Multiple gene expression; filamentous fungal cell; EST;
KW XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW XX culture condition; environmental stress; spore morphogenesis;
KW XX metabolic pathway engineering; catabolic pathway engineering; ss.
XX XX
OS XX Fusarium venenatum.
XX XX
PN XX W0200056762-A2.
XX XX
PD XX 28-SEP-2000.
XX XX
PF XX 22-MAR-2000; 2000WO-US07781.
XX XX
PR XX 22-MAR-1999; 99US-0273623.
XX XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX XX (NOVO ) NOVO NORDISK AS.
XX XX
PI XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX XX WPI; 2000-594572/56.
XX XX
XX XX Monitoring differential expression of genes in filamentous fungal cells
XX XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX XX substrate of expressed sequence tags -
XX XX
XX XX Claim 86; Page 1051; 3161pp; English.
XX XX
XX XX The present invention describes a method for monitoring differential
XX XX expression of genes in a first filamentous fungal (FF) cell relative to
XX XX expression of the same genes in one or more second filamentous fungal
XX XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX XX are used in the methods for monitoring differential expression of genes
XX XX in a first filamentous fungal (FF) cell relative to expression of the
XX XX same genes in one or more second filamentous fungal cells. Monitoring
XX XX the global expression of genes from FF cells allows the production
XX XX potential of the microorganisms to be improved. New genes may be
XX XX discovered, possible functions of unknown open reading frames can be
XX XX identified and gene copy number variation and stability can be
XX XX monitored. The expression of genes can be used to study how FF cells
XX XX adapt to changes in culture conditions, environmental stress, spore
XX XX morphogenesis, recombination, metabolic or catabolic pathway
XX XX engineering. Using ESTs provides several advantages over genomic or
XX XX random cDNA clones including elimination of redundancy as one spot on an
XX XX array equals one gene or open reading frame, and organisation of the
XX XX microarrays based on function of the gene products to facilitate
XX XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX XX Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX XX niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX XX AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX XX all specifically claimed in the present invention.
XX XX
SQ XX Sequence 629 BP; 153 A; 169 C; 135 G; 168 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 629;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatt 14
DB 83 catcgccagcatt 95
|||||
RESULT 28
AAZ17005
ID AAZ17005 standard; cDNA; 714 BP.
XX
XX AAZ17005;

```

```

XX DT 12-OCT-1999 (first entry)
XX XX
XX XX Human gene expression product cDNA sequence SEQ ID NO:4475.
XX XX
XX XX Human; gene: gene expression product; diagnosis; therapy; probe;
XX XX detection; mapping; tissue typing; profiling; forensic; cancer;
XX XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX XX
XX XX Homo sapiens.
XX XX
XX XX W09938972-A2.
XX XX
XX XX 05-AUG-1999.
XX XX
XX XX 28-JAN-1999; 99WO-US01619.
XX XX
XX XX 03-APR-1998; 98US-0080666.
XX XX
XX XX 28-JAN-1998; 98US-0072910.
XX XX
XX XX 24-FEB-1998; 98US-0075954.
XX XX
XX XX 31-MAR-1998; 98US-0080114.
XX XX
XX XX 03-APR-1998; 98US-0080515.
XX XX
XX XX (CHIR ) CHIRON CORP.
XX XX (HYSE ) HYSEQ INC.
XX XX
XX XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX XX
XX XX WPI; 1999-494092/41.
XX XX
XX XX Novel human genes and their expression products which are
XX XX differentially expressed in different cell types
XX XX
XX XX Claim 1; Page 2120; 2479pp; English.
XX XX
XX XX The present invention describes a library of human polynucleotides
XX XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX XX a method of detecting differentially expressed genes correlated with the
XX XX cancerous state of a mammalian cell, comprising detecting at least one
XX XX differentially expressed gene product in a test sample from a cell
XX XX suspected of being cancerous, where the gene product is encoded by one
XX XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX XX polynucleotides can be used as a source of primers and probes, which can
XX XX be used for a variety of purpose, e.g. detection of expression levels,
XX XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX XX can be used for raising antibodies for experimental, diagnostic and
XX XX therapeutic purposes. The polynucleotides may also be used to construct
XX XX arrays for diagnostics (which may be used to determine function of an
XX XX encoded protein); and to detect differences in expression levels between
XX XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX XX identify a genetic predisposition or susceptibility to a disease such as
XX XX cancer). The polynucleotides of the invention are especially used in the
XX XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX XX and lung cancer. The polynucleotides can also be used to screen for
XX XX peptide analogues and antagonists.
XX XX
XX XX Sequence 714 BP; 186 A; 142 C; 146 G; 231 T; 9 other;

Query Match 52.0%; Score 13; DB 20; Length 714;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcccagcattc 16
DB 74 tcgcccagcattc 86
|||||

```

```

RESULT 29
AAF22436
ID AAF22436 standard; cDNA; 725 BP.
XX
AC AAF22436;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:15.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX Obata Y;
XX
PI WPI; 2001-025274/03.
XX
DR
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
XX cancer.
XX
PS Claim 50; Page 276; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX
SQ Sequence 725 BP; 131 A; 225 C; 224 G; 136 T; 9 other;

Query Match 52.0%; Score 13; DB 22; Length 725;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
Db 229 gcatgccagcat 241
|||||

RESULT 30
AAZ17191/C
ID AAZ17191 standard; cDNA; 750 BP.
XX
AC AAZ17191;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4662.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

```

```

XX Homo sapiens.
OS
XX WO99398972-A2.
PN
XX
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99WO-US01619.
PF
XX
XX 03-APR-1998; 98US-0080666.
PR
XX 28-JAN-1998; 98US-0072910.
PR
XX 31-FEB-1998; 98US-0075954.
PR
XX 24-MAR-1998; 98US-0080114.
PR
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
PA
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
DR
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 2213-2214; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ1779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 750 BP; 176 A; 165 C; 156 G; 238 T; 15 other;

Query Match 52.0%; Score 13; DB 20; Length 750;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
Db 248 GCATGCCAGCAT 236
|||||

RESULT 31
AAC35170
ID AAC35170 standard; DNA; 974 BP.
XX
XX AAC35170;
XX
XX 17-OCT-2000 (first entry)
XX

```


PR 30-AUG-1999; 99US-01513303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 21-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 974;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 tttegatctcgt 25
 |||||

Db 593 tttegatctcgt 605

RESULT 32

AAC48947

ID AAC48947 standard; DNA; 974 BP.

XX

AC AAC48947;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59360.

XX

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 08-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.

Wed Oct 10 07:46:03 2001

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 974;
Best Local Similarity 100.0%; Pred. No. 1.le-02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ttctgatctctcgt 25
|||||
Db 593 ttctgatctctcgt 605

RESULT 33
AAC46920/c
ID AAC46920 standard; DNA; 985 BP.

XX AAC46920;

XX AC
XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51892.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123380.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.


```
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158493.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.08; Score 13; DB 21; Length 985;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 8 cagcatttcgac 20
    |||||
Db 938 CAGCATTTCGATC 926

RESULT 34
AAC47663/c
ID AAC47663 standard; DNA; 996 BP.
XX
AC AAC47663;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54655.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PP 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 30-APR-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
```

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 996;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ttctgatactcgt 25
|||||
Db 714 TTTCGATCCGCT 702

RESULT 35

AAZ96443
ID AAZ96443 standard; DNA; 1007 BP.

XX AC AAZ96443;

XX DT 10-APR-2000 (first entry)

XX DE S. pneumoniae derived DNA from ORF #271.

XX KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
XX KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.

XX OS Streptococcus pneumoniae.

XX PN WO9806734-A1.

XX PD 19-FEB-1998.

XX PF 15-AUG-1997; 97WO-US14436.

XX PR 16-AUG-1996; 96US-0024022.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Black Mt., Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;

XX WPI; 1998-159452/14.
XX P-PSDB; AAY86128.

XX PT Streptococcus pneumoniae proteins and related DNA - useful for
XX PT screening compounds for antibacterial activity

XX PS Claim 4; Page 287; 640pp; English.

CC This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AAF96173-96494) and their encoded proteins (see
 CC AAF8792-96618). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX
 SQ Sequence 1007 BP; 327 A; 206 C; 161 G; 313 T; 0 other;

Query Match 52.0%; Score 13; DB 19; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 Db 506 gcatgccagcat 518

RESULT 36
 AAF12052
 ID AAF12052 standard; cDNA; 1038 BP.
 AC AAF12052;
 XX
 XX 13-MAR-2001 (first entry)
 DT
 XX Aspergillus oryzae EST SEQ ID NO:4575.
 DE
 XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 XX WC200056762-A2.
 PN
 XX
 XX 28-SEP-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07781.
 XX
 XX 22-MAR-1999; 99US-0273623.
 XX
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO NORDISK AS.
 XX
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells
 XX uses fluorescence-labeled nucleic acids isolated from the cells and a
 XX substrate of expressed sequence tags -
 XX
 XX Claim 88; Page 1948; 3161pp; English.
 XX
 XX The present invention describes a method for monitoring differential
 XX expression of genes in a first filamentous fungal (FF) cell relative to
 XX expression of the same genes in one or more second filamentous fungal
 XX cells. The method uses fluorescence-labeled nucleic acids isolated from
 XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 XX are used in the methods for monitoring differential expression of genes
 XX in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

SQ Sequence 1038 BP; 258 A; 294 C; 243 G; 243 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1038;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 Db 593 gcatgccagcat 605

RESULT 37
 AAX84999
 ID AAX84999 standard; DNA; 1065 BP.
 XX
 AC AAX84999;
 XX
 XX 30-JUL-1999 (first entry)
 DT
 XX Human secreted protein gene No. 67.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.
 XX
 XX WO9924836-A1.
 XX
 XX 20-MAY-1999.
 XX
 XX 04-NOV-1998; 98WO-US23435.
 XX
 XX 17-NOV-1997; 97US-0066100.
 XX 07-NOV-1997; 97US-0064900.
 XX 07-NOV-1997; 97US-0064908.
 XX 07-NOV-1997; 97US-0064911.
 XX 07-NOV-1997; 97US-0064912.
 XX 07-NOV-1997; 97US-0064983.
 XX 07-NOV-1997; 97US-0064984.
 XX 07-NOV-1997; 97US-0064985.
 XX 07-NOV-1997; 97US-0064987.
 XX 07-NOV-1997; 97US-0064988.
 XX 17-NOV-1997; 97US-0066090.
 XX 17-NOV-1997; 97US-0066094.
 XX 17-NOV-1997; 97US-0066095.
 XX 17-NOV-1997; 97US-0066089.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Ebner R, Endress GA, Peng P, Janat F;
 PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Wei Y;
 XX WPI: 1999-337740/28.
 DR P-PSDB; AAY27633.
 XX
 PT New human secreted proteins and coding sequences useful for treating
 PT disorders of the immune system and hyperproliferative disorders
 XX
 PS Claim 1; Page 316; 507pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number is given in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAX84924) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 125 novel genes and their fragments (nucleic
 CC acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 125
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX84933 for described uses).
 XX
 SQ Sequence 1065 BP; 217 A; 297 C; 248 G; 303 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 Db 829 gcatgccagcat 841

RESULT 38
 AAT84197
 ID AAT84197 standard; DNA; 1110 BP.
 XX
 AC AAT84197;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE DNA encoding a UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase.
 XX
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome; ss.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 CDS 570..974
 FT /*tag= a
 FT
 XX WO9730070-A1.
 XX
 PD 21-AUG-1997.
 XX
 XX 19-FEB-1997; 97WO-US02318.
 XX
 XX 20-FEB-1996; 96US-0011888.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.
 DR P-PSDB; AAW28300.
 XX
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 XX
 PS Claim 9; Page 953; 989pp; English.
 XX
 CC The present sequence encodes a Staphylococcus aureus protein, that,
 CC based on homology with an E. coli protein, is believed to be a
 CC UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase (EC 1.1.1.-) (also
 CC known as UDP-mannac dehydrogenase). The present sequence was obtained
 CC from a library of clones of S. aureus WCUH 29 in E. coli. The DNA
 CC sequence can be used in the construction of ribozymes and antisense
 CC sequences to control the expression of staphylococcal genes. The DNA
 CC sequence is also useful as a source of regulatory elements for the
 CC control of bacterial gene expression. The encoded protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 XX
 SQ Sequence 1110 BP; 385 A; 152 C; 222 G; 334 T; 17 other;

Query Match 52.0%; Score 13; DB 18; Length 1110;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagcattt 15
 |||||
 Db 680 atcgccagcattt 692

RESULT 39
 AAC41953/C
 ID AAC41953 standard; DNA; 1114 BP.
 XX
 AC AAC41953;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33748.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145195.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1114;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ttctgattctctgt 25

Db 674 TTTCGATCTCTGT 662

RESULT 40

AAAL15961

ID AAAL15961 standard; cDNA; 1179 BP.

XX AC AAAL15961;

DT 12-JUN-2000 (first entry)

DE Human protein clone HP10530 coding sequence.

XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.

XX Homo sapiens.

OS WO200005367-A2.

PN 03-FEB-2000.

XX 22-JUL-1999; 99WO-JP03929.

XX 24-JUL-1998; 98JP-0208820.

PR 07-AUG-1998; 98JP-0224105.

PR 25-AUG-1998; 98JP-0238116.

PR 09-SEP-1998; 98JP-0254736.

PR 29-SEP-1998; 98JP-0275505.

XX (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX Kato S, Kimura T;
 PI WPI; 2000-182694/16.
 DR P-PSDB; AAY9483.

XX

PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX Claim 3; Page 289-290; 351pp; English.

XX This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.

XX Sequence 1179 BP; 244 A; 343 C; 356 G; 236 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1179;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcattctgat 19

Db 549 ccagcattctgat 561

RESULT 41

AAAC36239/C

ID AAC36239 standard; DNA; 1226 BP.

XX AC AAC36239;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13057.

XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PD

XX	25-FEB-1999;	99US-0121825.	PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.		PR	09-JUL-1999;	99US-0142920.
XX			PR	12-JUL-1999;	99US-0142977.
XX	25-FEB-1999;	99US-0121825.	PR	13-JUL-1999;	99US-0143542.
PR	05-MAR-1999;	99US-0123180.	PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134236.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.

```

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1226;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgat 19
Db 591 CCAGCATTTCGAT 579

RESULT 42
ID AAN82205 standard; DNA; 1263 BP.
AC AAN82205;
XX
XX
DT 12-NOV-1990 (first entry)
DE
DE Nad B gene encoding L aspartate 3' end.
XX
XX Quinolinic acid synthase; L aspartate oxidase; nad A; nad B;
KW plant protection agent; ss.
XX
XX Escherichia coli (K12C600).
XX
XX DE3703255-A.
XX
XX 18-AUG-1988.
XX
XX 04-FEB-1987; 87DE-3703255.
XX
XX 04-FEB-1987; 87DE-3703255.
XX
XX (RUTG ) RURGERSWERKE AG.
XX
XX Laufer A, Gasßen HG, Flachmann R, Hoke H, Holtmann W, Kunz N;
PI

```

```

PI Stadelhofer J, Seifert J;
XX
XX WPI; 1988-235937/34.
XX
XX Plasmids coded for quinolinic acid synthesis - have DNA sequences contg.
PT genetic information for synthesis of quinolinic synthase and L aspartate
PT oxidase.
XX
XX Claim 7; Page 4-5; 24pp; German.
XX
XX Between this 3' end sequence and the 5' end sequence of AAN80036 are
CC ca. 200 bp not represented in the specification.
CC The nad A sequence (AAN80035) encoding a polypeptide with the biological
CC activity of quinolinic acid synthase, together with nad B encoding a
CC polypeptide with the biological activity of L aspartate is inserted into
CC a plasmid. The quinolinic acid expressed is an intermediate for
CC pharmaceuticals and plant protection agents.
XX
XX Sequence 1263 BP; 307 A; 324 C; 344 G; 284 T; 4 other;
SQ

Query Match 52.0%; Score 13; DB 9; Length 1263;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgccagcatttc 16
Db 658 tcgccagcatttc 670

RESULT 43
ID AAA93673 standard; DNA; 1287 BP.
AC AAA93673;
XX
XX 16-JAN-2001 (first entry)
XX
XX Human peroxisome associated protein splice variant SECX 3884846-1 DNA.
DE
DE SECX protein; human; secreted; membrane-associated; cancer;
XX proliferation regulator; differentiation regulator; non-malignant tumour;
XX immune disorder; autoimmune disease; transplant rejection; allergy; AIDS;
XX infection; inflammatory disorder; arthritis; haematopoietic disorder;
XX skin disorder; cardiovascular disorder; atherosclerosis; restenosis;
XX neurological disease; Alzheimer's disease; trauma; wound;
XX spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;
XX anti-HIV; antiinflammatory; antiarthritic; antiarteriosclerotic;
XX neuroprotective; vulnery; antiallergic; antimicrobial; cardiant;
XX dermatological; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200053742-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06280.
XX
XX 09-MAR-1999; 99US-0123667.
XX 08-MAR-2000; 2000US-0123667.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA;
PI
XX
XX WPI; 2000-594318/56.
XX P-PSDB; AAB23045.
XX
XX Novel human membrane associated or secreted polypeptides and
PT polynucleotides useful for diagnosis, prevention and treatment of
PT pathological states such as cancer, immune, cardiovascular and
PT neurological disorders

```


XX PS Claim 3; Fig 7b; 151pp; English.

XX CC Sequences AAA93616-A93631 and AAA93673-A93676 represent nucleic acids

XX CC which encode human SECX proteins (AAB23029-B23048). The SECX proteins

XX CC of the invention are either secreted or membrane-associated proteins

XX CC and act as regulator of cellular proliferation and differentiation. SECX

XX CC proteins or nucleotides are useful for diagnosing the presence of, or

XX CC predisposition to, a disease associated with altered levels of SECX

XX CC proteins and nucleotides. The SECX proteins are also useful to screen

XX CC compounds that modulate SECX activity or expression. The interaction of

XX CC a SECX protein with other cellular proteins may be useful to modulate

XX CC the activity of a partner protein, cellular proliferation, cellular

XX CC differentiation and cell survival. SECX nucleotides are useful for the

XX CC recombinant expression of SECX protein, and may be used to detect SECX mRNA

XX CC or genetic lesions in the SECX gene. They may also be used to modulate

XX CC SECX expression (e.g., using antisense oligonucleotides). SECX nucleic

XX CC acid sequences are also useful for identifying a cell or tissue type in

XX CC a biological sample, and in forensic biology. SECX primers or probes are

XX CC useful for detecting the presence of SECX nucleotides and for screening

XX CC tissue cultures for contamination. Diseases that may be treated or

XX CC prevented using SECX proteins or nucleotides include cancer (e.g.,

XX CC colorectal carcinoma, prostate cancer), benign tumours, immune disorders

XX CC (including autoimmune diseases, transplant rejection, allergies, AIDS),

XX CC infections, inflammatory disorders, arthritis, haematopoietic disorders,

XX CC skin disorders, cardiovascular disorders, atherosclerosis, restenosis,

XX CC neurological diseases (e.g., Alzheimer's disease), trauma (e.g.,

XX CC surgical or traumatic wounds, spinal cord injury), and skeletal

XX CC disorders.

XX SQ Sequence 1287 BP; 266 A; 368 C; 388 G; 262 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 1287;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgat 19

Db 609 ccagcattcgat 621

|||||

RESULT 44

AAA93674

ID AAA93674 standard; DNA; 1291 BP.

XX AC AAA93674;

XX DT 16-JAN-2001 (first entry)

XX DE Human peroxisome associated protein splice variant SECX 3884846-2 DNA.

XX KW SECX protein; human; secreted; membrane-associated; cancer;

XX KW proliferation regulator; differentiation regulator; non-malignant tumour;

XX KW immune disorder; autoimmune disease; transplant rejection; allergy; AIDS;

XX KW infection; inflammatory disorder; arthritis; haematopoietic disorder;

XX KW skin disorder; cardiovascular disorder; atherosclerosis; restenosis;

XX KW neurological disease; Alzheimer's disease; trauma; woundings;

XX KW spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;

XX KW anti-HIV; antiinflammatory; antiarthritic; antiarteriosclerotic;

XX KW neuroprotective; vulnary; antiallergic; antimicrobial; cardiant;

XX KW dermatological; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200053742-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06280.

XX PR 09-MAR-1999; 99US-0123667.

XX PR 08-MAR-2000; 2000US-0123667.

(CURA-) CURAGEN CORP.

Shinkets RA;

WPI; 2000-594318/56.

P-PSDB; AAB23046.

Novel human membrane associated or secreted polypeptides and

polynucleotides useful for diagnosis, prevention and treatment of

pathological states such as cancer, immune, cardiovascular and

neurological disorders

Claim 3; Fig 7C; 151pp; English.

Sequences AAA93616-A93631 and AAA93673-A93676 represent nucleic acids

which encode human SECX proteins (AAB23029-B23048). The SECX proteins

of the invention are either secreted or membrane-associated proteins

and act as regulator of cellular proliferation and differentiation. SECX

proteins or nucleotides are useful for diagnosing the presence of, or

predisposition to, a disease associated with altered levels of SECX

proteins and nucleotides. The SECX proteins are also useful to screen

compounds that modulate SECX activity or expression. The interaction of

a SECX protein with other cellular proteins may be useful to modulate

the activity of a partner protein, cellular proliferation, cellular

differentiation and cell survival. SECX nucleotides are useful for the

recombinant expression of SECX protein, and may be used to detect SECX mRNA

or genetic lesions in the SECX gene. They may also be used to modulate

SECX expression (e.g., using antisense oligonucleotides). SECX nucleic

acid sequences are also useful for identifying a cell or tissue type in

a biological sample, and in forensic biology. SECX primers or probes are

useful for detecting the presence of SECX nucleotides and for screening

tissue cultures for contamination. Diseases that may be treated or

prevented using SECX proteins or nucleotides include cancer (e.g.,

colorectal carcinoma, prostate cancer), benign tumours, immune disorders

(including autoimmune diseases, transplant rejection, allergies, AIDS),

infections, inflammatory disorders, arthritis, haematopoietic disorders,

skin disorders, cardiovascular disorders, atherosclerosis, restenosis,

neurological diseases (e.g., Alzheimer's disease), trauma (e.g.,

surgical or traumatic wounds, spinal cord injury), and skeletal

disorders.

Sequence 1291 BP; 267 A; 370 C; 389 G; 262 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 1291;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgat 19

Db 613 ccagcattcgat 625

|||||

RESULT 45

AAA61879

ID AAC61879 standard; cDNA; 1313 BP.

XX AC AAC61879;

XX XX 06-MAR-2001 (first entry)

XX DE cDNA encoding a human secreted protein.

XX KW Secreted protein; platelet disorder; stem cell disorder; osteoporosis;

XX KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;

XX KW nervous system disease; bone growth; cosmetic plastic surgery;

XX KW gut protection; gut regeneration; fibrosis; cancer;

XX KW bone marrow transplantation; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

```

FT CDS      163..1140
FT          /*tag= a
FT          /product= "secreted protein"
XX PN       WO2000061755-A2.
XX PD       19-OCT-2000.
XX PF       10-APR-2000; 2000WO-US09555.
XX PR       09-APR-1999; 99US-0128574.
XX PR       20-AUG-1999; 99US-0150054.
XX PA       (CHIR ) CHIRON CORP.
XX PI       Garcia PD;
XX DR       WPI; 2000-665133/64.
XX DR       P-PSDB; AAB19393.
XX PT       Novel secreted human proteins useful for stimulating blood cell
XX PT       generation in patients receiving cancer chemotherapy, treating bone
XX PT       marrow transplantation patients and for healing fractured bones -
XX PS       Claim 2; Page 54; 74pp; English.
XX CC       AAC61879-93 encode secreted human proteins. The secreted proteins are
XX CC       useful in assays to determine their biological activities. The proteins
XX CC       can also be used as biomarkers to identify tissues or cell types which
XX CC       express the proteins. The polynucleotide molecules can be used as
XX CC       biomarkers for tissues or chromosomes and to elicit immune responses.
XX CC       The proteins and antibodies are useful in diagnosis and treatment of
XX CC       diseases associated with altered expression of these proteins. The
XX CC       proteins are also useful for prevention or treatment of platelet
XX CC       disorders, stem cell disorders, osteoporosis or osteoarthritis, burns,
XX CC       incisions, ulcers, periodontal diseases, central and peripheral nervous
XX CC       system diseases and neuropathies, for healing fractured bones and to
XX CC       induce cartilage and/or bone growth in cosmetic plastic surgery. The
XX CC       proteins are also useful for gut protection or regeneration. The
XX CC       treatment of lung or liver fibrosis, for stimulating blood cell
XX CC       generation in patients receiving cancer chemotherapy and for treatment
XX CC       of bone marrow transplantation patients.
XX SQ       Sequence 1313 BP; 298 A; 366 C; 360 G; 289 T; 0 other;

Query Match      52.0%; Score 13; DB 21; Length 1313;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ccagcatttcgat 19
Db      711 ccagcatttcgat 723

Search completed: October 9, 2001, 15:52:46
Job time: 15127 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:36 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-7

Perfect score: 25

Sequence: 1 gcacgcagcattgcattcctcgt 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8423

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	2232	5	PCT-US96-05320A-113
6	14	56.0	5319	1	US-08-169-927-1
7	13	52.0	61	4	US-09-023-228B-59
8	13	52.0	447	4	US-08-836-075A-51
9	13	52.0	1460	2	US-08-933-750C-60
10	13	52.0	1460	3	US-09-234-613-60
11	13	52.0	1850	2	US-08-743-637B-172
12	13	52.0	1650	3	US-08-526-840B-172
13	13	52.0	2104	1	US-08-682-193A-1
14	13	52.0	2193	1	US-08-420-235B-16
15	13	52.0	2193	3	US-08-793-624-16
16	13	52.0	2193	5	PCT-US95-10194-16
17	13	52.0	2312	2	US-09-134-566-15
18	13	52.0	2314	2	US-09-134-566-8
19	13	52.0	2973	2	US-08-343-101A-5
20	13	52.0	2973	3	US-09-183-688-5
21	13	52.0	20710	1	US-08-420-235B-1
22	13	52.0	20710	3	US-08-793-624-1
23	13	52.0	20710	5	PCT-US95-10194-1
24	13	52.0	35100	2	US-08-770-379-18
25	13	52.0	35100	4	US-08-757-669A-18
26	12	48.0	19	3	US-08-851-843A-92
27	12	48.0	19	4	US-08-974-549A-384

c 28 12 48.0 38 1 US-08-373-124A-1692 Sequence 1692, Ap
c 29 12 48.0 38 1 US-08-435-628-1692 Sequence 1692, Ap
c 30 12 48.0 660 2 US-08-743-637B-175 Sequence 175, App
c 31 12 48.0 660 3 US-08-526-840B-175 Sequence 175, App
32 12 48.0 715 4 US-08-998-416-767 Sequence 767, App
33 12 48.0 780 3 US-08-927-128-3 Sequence 3, Appli
34 12 48.0 841 4 US-08-990-823-39 Sequence 39, Appli
35 12 48.0 864 2 US-08-595-868C-11 Sequence 11, Appli
36 12 48.0 867 4 US-09-139-819A-11 Sequence 11, Appli
c 37 12 48.0 954 1 US-08-624-125-9 Sequence 9, Appli
c 38 12 48.0 1188 1 US-08-781-562-2 Sequence 2, Appli
c 39 12 48.0 1239 4 US-09-064-693A-20 Sequence 20, Appli
40 12 48.0 1249 3 US-08-985-950-1 Sequence 1, Appli
c 41 12 48.0 1305 2 US-08-484-575A-19 Sequence 19, Appli
c 42 12 48.0 1305 3 US-08-477-459-19 Sequence 19, Appli
c 43 12 48.0 1305 3 US-08-479-869-19 Sequence 19, Appli
c 44 12 48.0 1305 3 US-08-486-414-19 Sequence 19, Appli
c 45 12 48.0 1305 5 PCT-US94-01826A-19 Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA: US/08401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcatttcgattcctcgt 25
|||||
Db 90 GCATGCCAGCAGCATTCGATCCTCGT 114

RESULT 2

US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcatttcgattcctcgt 25
|||||
Db 90 GCATGCCAGCAGCATTCGATCCTCGT 114

RESULT 3

US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioB"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: aminotransf."
FEATURE:
NAME/KEY: CDS

```

; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORF1"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORF1 RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-1

```

```

Query Match      100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 8.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcatgcgcagcatttcgattcctcgt 25
   |||||||
DB 206 GCATGCCGCGAGATTTCGATPCCTCGT 230

```

RESULT 4

```

US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

```

```

; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioF"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "DTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1

```


PUBLICATION INFORMATION:

AUTHORS: Carl, M.
AUTHORS: Dobson, M. E.
AUTHORS: Ching, W. M.
AUTHORS: Dasch, G. A.
TITLE: Characterization of the gene encoding the
TITLE: protective S-layer protein of Rickettsia
TITLE: prowazekii; presence of a truncated identical
TITLE: homolog in rickettsia typhi
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattt 15
|||||
Db 2221 CATGCCAGCATT 2208

RESULT 7

US-09-023-228B-59
; Sequence 59, Application US/09023228B
; Patent No. 6140490
; GENERAL INFORMATION:

APPLICANT: BIESECKER, GREGORY
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS
NUMBER OF SEQUENCES: 157
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Place #200
CITY: Denver
STATE: Colorado
COUNTRY: US

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,228B
FILING DATE: 12-FEBRUARY-1998
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01739
FILING DATE: 30 JAN 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/595,335

FILING DATE: 1 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson, Esq.
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX50/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:

OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-59

Query Match 52.0%; Score 13; DB 3; Length 61;
Best Local Similarity 69.2%; Pred. No. 46;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatc 20
|||||
Db 23 CAGCAUUCGAUC 35

RESULT 8

US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995

PRIOR APPLICATION DATA: EP 94870166.9
APPLICATION NUMBER: 21 Oct 1994
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:

LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatt 14
|||||
Db 406 CATGCCAGCATT 418

RESULT 9

US-08-933-750C-60
; Sequence 60, Application US/08933750C

```

; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOR01
; CLONE: 140516
; US-09-933-750C-60

```

```

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgat 19
Db 722 CCAGCATTTCGAT 734

```

```

RESULT 10
US-09-234-613-60
; Sequence 60, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.

```

```

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOR01
; CLONE: 140516
; US-09-234-613-60

```

```

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgat 19
Db 722 CCAGCATTTCGAT 734

```

```

RESULT 11
US-08-743-637B-172/c
; Sequence 172, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 172:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-743-637B-172

Query Match 52.0%; Score 13; DB 2; Length 1650;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatc 20
| | | | | | | | | |
Db 555 CAGCATTTCGATC 543

RESULT 12
US-08-526-840B-172/c
; Sequence 172, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: QUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: APPLICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526.840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
```

```
;
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 172:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-526-840B-172

Query Match 52.0%; Score 13; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatc 20
| | | | | | | | | |
Db 555 CAGCATTTCGATC 543

RESULT 13
US-08-682-193A-1
; Sequence 1, Application us/08682193A
; Patent No. 5776740
; GENERAL INFORMATION:
; APPLICANT: HATAKEYAMA, Kazuhisa
; APPLICANT: GOTO, Makoto
; APPLICANT: TERASAWA, Masato
; APPLICANT: YUKAWA, Hideaki
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF
; TITLE OF INVENTION: L-TRYPTOPHANE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02019
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,193A
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7-181730
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: DAVID, RESNICK S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 46643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: 200291 STRE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Brevibacterium flavum
; STRAIN: MJ-233
```

;
; NAME/KEY: Coding Sequence
; LOCATION: 556...1855
; OTHER INFORMATION:
US-08-682-193A-1

Query Match 52.0%; Score 13; DB 1; Length 2104;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 ttgcgacctcgt 25
Db 1635 TTTCGATCCTCGT 1647

RESULT 14
US-08-420-235B-16/C
; Sequence 16, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:

NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
US-08-420-235B-16

Query Match 52.0%; Score 13; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
Db 45 GCATGCCAGCAT 33

RESULT 15
US-08-793-624-16/c
; Sequence 16, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.

; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2193
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 52.0%; Score 13; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
Db 45 GCATGCCAGCAT 33

RESULT 16
PCT-US95-10194-16/C
; Sequence 16, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MS
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:

NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match 52.0%; Score 13; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcat 13
|||||
Db 45 GCATGCCAGCAT 33

RESULT 17
US-09-134-566-15
; Sequence 15, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; EARLIER FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-15

Query Match 52.0%; Score 13; DB 2; Length 2312;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcat 16
|||||
Db 960 tcgccagcat 972

RESULT 18
US-09-134-566-8
; Sequence 8, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; EARLIER FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8
LENGTH: 2314
TYPE: DNA
ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcat 16
|||||
Db 962 tcgccagcat 974

RESULT 19
US-08-343-101A-5/c
; Sequence 5, Application US/08343101A
; Patent No. 5830759
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,101A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-343-101A-5

Query Match 52.0%; Score 13; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcat 13
|||||
Db 2181 GCATGCCAGCAT 2169

RESULT 20
US-09-183-688-5/c
; Sequence 5, Application US/09183688

```

; Patent No. 6093550
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,688
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/343,101
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-09-183-688-5

```

```

Query Match          52.0%; Score 13; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 gcatcgccagcat 13
   |||||
Db 2181 GCATCGCCAGCAT 2169

```

```

RESULT 21
US-08-420-235B-1/c
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-420-235B-1

```

```

Query Match          52.0%; Score 13; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 gcatcgccagcat 13
   |||||
Db 2137 GCATCGCCAGCAT 2125

```

```

RESULT 22
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

```

```

Query Match          52.0%; Score 13; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 gcatcgccagcat 13
   |||||
Db 2137 GCATCGCCAGCAT 2125

```

```

RESULT 23
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF

```

```
;
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10194
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; PCT-US95-10194-1

Query Match 52.0%; Score 13; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcgcagcat 13
|||||
Db 2137 GCATGCCAGCAT 2125

RESULT 24
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
;
```

```
;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-18

Query Match 52.0%; Score 13; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcgcagcat 13
|||||
Db 2057 GCATGCCAGCAT 2045

RESULT 25
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
;
```

Query Match 52.0%; Score 13; DB 4; Length 35100;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 DB 2057 GCATGCCAGCAT 2045

RESULT 26
 US-08-851-843A-92/c
 ; Sequence 92, Application US/08851843A
 ; Patent No. 6093809
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.
 ; APPLICANT: Harley, Calvin
 ; APPLICANT: Andrews, William H.
 ; TITLE OF INVENTION: No. 6093809el Telomerase
 ; NUMBER OF SEQUENCES: 225
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/851,843A
 ; FILING DATE: 06-MAY-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/846,017
 ; FILING DATE: 25-APR-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/844,419
 ; FILING DATE: 18-APR-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 01-OCT-1996
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Apple, Randolph T.
 ; REGISTRATION NUMBER: 36,429
 ; REFERENCE/DOCKET NUMBER: 015389-00293005
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 92:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcat 13

DB 19 CATGCCAGCAT 8

RESULT 27
 US-08-974-549A-384/c
 ; Sequence 384, Application US/08974549A
 ; Patent No. 6166178
 ; GENERAL INFORMATION:
 ; APPLICANT: Cech, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.
 ; APPLICANT: Harley, Calvin B.
 ; APPLICANT: Andrews, William H.
 ; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 ; NUMBER OF SEQUENCES: 727
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/974,549A
 ; FILING DATE: 19-NOV-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/724,643
 ; FILING DATE: 01-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/844,419
 ; FILING DATE: 18-APR-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/846,017
 ; FILING DATE: 25-APR-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/851,843
 ; FILING DATE: 06-MAY-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/854,050
 ; FILING DATE: 09-MAY-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/911,312
 ; FILING DATE: 14-AUG-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/912,951
 ; FILING DATE: 14-AUG-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/915,503
 ; FILING DATE: 14-AUG-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/US97/17618
 ; FILING DATE: 01-OCT-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/US97/17885
 ; FILING DATE: 01-OCT-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Apple, Randolph Ted
 ; REGISTRATION NUMBER: 36,429
 ; REFERENCE/DOCKET NUMBER: 015389-0026100S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 384:
 ; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 19 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..19
; OTHER INFORMATION: /note= "Tcpl.6 primer"
US-08-974-549A-384

Query Match 48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcat 13
Db 19 CATGCCAGCAT 8
|||||

RESULT 28
US-08-373-124A-1692/c
; Sequence 1692, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1692:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

;
; TOPOLOGY: linear
US-08-373-124A-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gccagcatttcg 17
Db 38 GCCAGCATTTTCG 27
|||||

RESULT 29
US-08-435-628-1692/c
; Sequence 1692, Application US/08435628
; Patent No. 5817796
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,628
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,124
; FILING DATE: January 13, 1995
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1692:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-435-628-1692

Query Match

48.0%; Score 12; DB 1; Length 38;

```

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gccagcattcg 17
   |||||
Db 38 GCCAGCATTCG 27

RESULT 30
US-08-743-637B-175/c
; Sequence 175, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743.637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-743-637B-175

Query Match 48.0%; Score 12; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgat 19
   |||||
Db 627 CAGCATTCGAT 616

RESULT 31
US-08-840B-175/c
; Sequence 175, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.

```

```

; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-526-840B-175

Query Match 48.0%; Score 12; DB 3; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgat 19
   |||||
Db 627 CAGCATTCGAT 616

RESULT 32
US-08-998-416-767
; Sequence 767, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA

```


ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meligs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 767:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1495RP
US-08-998-416-767

Query Match 48.0%; Score 12; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcattt 15
Db 356 TCGCCAGCATTT 367

RESULT 33
US-08-927-128-3
Sequence 3, Application US/08927128
Patent No. 6127150
GENERAL INFORMATION:
APPLICANT: Coolidge, Thomas
APPLICANT: Wagner, Fred
APPLICANT: ven Heeke, Gino
APPLICANT: Schuster, Sheldon
APPLICANT: Stout, Jay
APPLICANT: Wylie, Dwane
TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchants & Gould
STREET: 3100 No. 6127150 West Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,128
FILING DATE: 05-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/680,004
FILING DATE: 15-JUL-1995

ATTORNEY/AGENT INFORMATION:
NAME: Carter, Charles G
REGISTRATION NUMBER: 35,093
REFERENCE/DOCKET NUMBER: 8648.2USDI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..780
OTHER INFORMATION:
US-08-927-128-3

Query Match 48.0%; Score 12; DB 3; Length 780;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ttgcgtctcgt 25
Db 532 TTGCATCTCTGT 543

RESULT 34
US-08-990-823-39
Sequence 39, Application US/08990823D
Patent No. 6228371
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990,823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 39
LENGTH: 841
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-39

Query Match 48.0%; Score 12; DB 4; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcctcgccagca 12
Db 370 gcctcgccagca 381

RESULT 35
US-08-595-868C-11

```

; Sequence 11, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.590S01
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..864
; OTHER INFORMATION:
; US-08-595-868C-11

```

```

Query Match 48.0%; Score 12; DB 2; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 14 ttcgatctctgt 25
| | | | | | | | | |
Db 532 TTCGATCTCTGT 543

```

```

RESULT 36
US-09-139-819A-11
; Sequence 11, Application US/09139819A
; Patent No. 6251635
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.

```

```

; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
; TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,819A
; FILING DATE: 25-AUG-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/595,868
; FILING DATE: 06-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 089187/0144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..864
; US-09-139-819A-11

```

```

Query Match 48.0%; Score 12; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 14 ttcgatctctgt 25
| | | | | | | | | |
Db 532 TTCGATCTCTGT 543

```

```

RESULT 37
US-08-624-125-9/C
; Sequence 9, Application US/08624125
; Patent No. 5744341
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
; TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

```

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/624,125
;; FILING DATE: 29-MAR-1996
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KELBER, STEVEN B.
;; REGISTRATION NUMBER: 30,073
;; REFERENCE/DOCKET NUMBER: 2747-063-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-413-3000
;; TELEFAX: 703-413-2220
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 954 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-624-125-9

Query Match 48.0%; Score 12; DB 1; Length 954;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatt 13
|||||
Db 663 CATGCCAGCAT 652

RESULT 38
US-08-781-562-2/c
; Sequence 2, Application US/08/81562
; Patent No. 5763589
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,562
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0181 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1188 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: Consensus
;; CLONE: Consensus
US-08-781-562-2

Query Match 48.0%; Score 12; DB 1; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cgcgcagcatttc 16
|||||
Db 72 CGCCAGCATTTTC 61

RESULT 39
US-09-064-693A-20/c
; Sequence 20, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-20

Query Match 48.0%; Score 12; DB 4; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgcgcagcatt 14
|||||
Db 1137 ATCGCCAGCAT 1126

```

RESULT 40
US-08-985-950-1
; Sequence 1, Application US/08985950
; Patent No. 6140076
; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 154..1062
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 211..1062
US-08-985-950-1

```

```

Query Match 48.0%; Score 12; DB 3; Length 1249;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY - 5 cgcacgatttc 16
      |||
Db 197 CGCCAGCATTTC 208

```

```

RESULT 41
US-08-484-575A-19/c
; Sequence 19, Application US/08484575A
; Patent No. 5925358
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran and David E. Junker

```

```

; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,575A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0450
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
US-08-484-575A-19

```

```

Query Match 48.0%; Score 12; DB 2; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CCAGCATTTCGA 18
      |||
Db 780 CCAGCATTTCGA 769

```

```

RESULT 42
US-08-477-459-19/c
; Sequence 19, Application US/08477459
; Patent No. 6001369
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,459
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

```

```
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
US-08-477-459-19

Query Match 48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
|||||
Db 780 CCAGCATTTTCGA 769

RESULT 43
US-08-479-869-19/c
; Sequence 19, Application US/08479869
; Patent No. 6123949
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D, Mark D
; TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and
; THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,869
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/024,156
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS

; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 1..1305
US-08-479-869-19

Query Match 48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
|||||
Db 780 CCAGCATTTTCGA 769

RESULT 44
US-08-486-414-19/c
; Sequence 19, Application US/08486414B
; Patent No. 6136318
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
; FILE REFERENCE: 42771D
; CURRENT APPLICATION NUMBER: US/08/486,414B
; CURRENT FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Fowlpox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(432)
US-08-486-414-19

Query Match 48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
|||||
Db 780 CCAGCATTTTCGA 769

RESULT 45
PCT-US94-01826A-19/c
; Sequence 19, Application PC/TUS9401826A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01826A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
```

```

; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
; PCT-US94-01826A-19

```

```

Query Match      48.0%; Score 12; DB 5; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ccagcatttcga 18
        |||
Db      780 CCAGCATTTCGA 769

```

Search completed: October 9, 2001, 15:55:39
 Job time: 13215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:32 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: us-09-396-196f-7
Perfect score: 25
Sequence: 1 gcatgccagatttcgatccctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 367121

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est33.*
111: gb_est34.*
112: gb_est35.*
113: gb_est36.*
114: gb_est37.*
115: gb_est38.*
116: gb_est39.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	68.0	498	21	AI532390 SD03891.5
2	16	64.0	429	107	AU030894 AU030894
3	16	64.0	510	107	AU031475 AU031475
4	16	64.0	666	107	AU030895 AU030895
5	16	64.0	704	107	AU078750 AU078750
6	15	60.0	258	246	AZ578088 AU078750
7	15	60.0	338	162	BE065518 RC3-BT030
8	15	60.0	341	31	AV551712 AV551712
9	15	60.0	393	158	H60965 YR22C07.r1
10	15	60.0	394	189	T86053 YR22C07.r1
11	15	60.0	394	190	W23358 SWACA507SK
12	15	60.0	399	162	BE044431 HO45D04.x
13	15	60.0	414	18	AI268648 Q039F05.x
14	15	60.0	435	223	AQ027122 CIT-HSP-2
15	15	60.0	437	158	H69200 YR95F10.r1
16	15	60.0	438	187	N92081 ZA33H03.r1
17	15	60.0	445	32	AV722315 AV722315
18	15	60.0	466	1	AA026650 ZJ99E08.r
19	15	60.0	475	158	H60979 YR22F01.r1
20	15	60.0	491	23	AI692369 W63F02.x
21	15	60.0	491	146	BF211097 G01812706
22	15	60.0	492	251	AZ924681 A906.1d55
23	15	60.0	507	188	T84279 YD47B04.r1
24	15	60.0	528	158	H74120 YS15E08.r1
25	15	60.0	539	170	BF839030 PM2-HT035
26	15	60.0	547	118	AW588900 RA06G05.Y
27	15	60.0	581	7	AA405993 ZU66F09.r
28	15	60.0	622	137	BE557935 F115A06.Y
29	15	60.0	630	137	BE583417 I1-6H-MY
30	15	60.0	684	145	BF156891 F158B12.Y
31	15	60.0	1096	145	BF144032 G01791288
32	14	56.0	195	189	W06020 TGESTZ87F1
33	14	56.0	200	251	AZ922752 SLC0F4A05
34	14	56.0	203	111	AW065325 G14045H11
35	14	56.0	203	155	BG589383 EST497225
36	14	56.0	205	155	BG558891 RH12-57
37	14	56.0	211	235	AQ931496 RPCI-23-2
38	14	56.0	231	18	AI329089 a9g12ne.r
39	14	56.0	242	103	AI902416 CM-BT006-
40	14	56.0	251	125	BB090429 BB090429
41	14	56.0	271	115	AA165335 ZQ49C06.r
42	14	56.0	286	3	AA165335
43	14	56.0	311	128	BB216989 BB216989
44	14	56.0	322	152	BG354207 947032F05
45	14	56.0	322	168	BF714260 mab35f04.

ALIGNMENTS

RESULT	1
AI532390/c	
LOCUS	
DEFINITION	AI532390.1 498 bp mRNA EST 18-MAR-1999
SD03891.5prime SD Drosophila melanogaster Schneider L2 cell culture	
POT2 Drosophila melanogaster cDNA clone SD03891 5prime, mRNA	
sequence.	
ACCESSION	AI532390.1 GI:4446525
VERSION	EST.
KEYWORDS	fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 498)
AUTHORS	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
	,P., Lewis,S. and Rubin,G.M.

TITLE
JOURNAL
COMMENT

BDGP/HHMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>
Plate: 38 row: H column: 7
High quality sequence stop: 433.

FEATURES
source

Location/Qualifiers
1..498
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SP03891"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell culture POT2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 158 a 106 c 124 g 110 t
ORIGIN

Query Match 68.0% Score 17: DB 21: Length 498;
Best Local Similarity 100.0%; Pred.No.2.5; 0; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 gcatcgccagcatttcg 17
|||||
Db 266 GCATCGCCAGCATTCG 250

RESULT 2

LOCUS AU030894/c 429 bp mRNA EST 29-OCT-1998
DEFINITION AU030894 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E60405_1A, mRNA sequence.
ACCESSION AU030894
VERSION AU030894.1 GI:3766784
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 (bases 1 to 429)
Sasaki,T. and Yamamoto,K.
Rice cDNA from immature leaf including apical meristem
Unpublished (1997)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305

FEATURES
source

Location/Qualifiers
1..429
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E60405_1A"
/clone_lib="Rice cDNA from immature leaf including apical meristem"
/dev_stage="immature"
/note="Organ: leaf; immature leaf including apical meristem (under long day condition)"
BASE COUNT 106 a 122 c 101 g 99 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 429;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 302 CATGCCAGCATTCG 287

RESULT 3

AU031475/c
 LOCUS AU031475 510 bp mRNA EST 30-OCT-1998
 DEFINITION AU031475 Rice cDNA from immature leaf including apical meristem
 ORYZA SATIVA cDNA clone E61659_22, mRNA sequence.

ACCESSION AU031475
 VERSION AU031475.1 GI:3767365
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1 (bases 1 to 510)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 Unpublished (1997)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = "RGP".

FEATURES

source
 1..510
 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone_lib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="immature"
 /note="Organ: leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 176 a 98 c 99 g 137 t

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 510;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 136 CATGCCAGCATTCG 121

RESULT 4

AU030895/c
 LOCUS AU030895 666 bp mRNA EST 29-OCT-1998
 DEFINITION AU030895 Rice cDNA from immature leaf including apical meristem
 ORYZA SATIVA cDNA clone E60405_22, mRNA sequence.

ACCESSION AU030895
 VERSION AU030895.1 GI:3766785
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1 (bases 1 to 666)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 Unpublished (1997)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = "RGP".

FEATURES
 source

1..666
 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone_lib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="immature"
 /note="Organ: leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 195 a 149 c 131 g 181 t 10 others

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 666;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 293 CATGCCAGCATTCG 278

RESULT 5

AU078750/c
 LOCUS AU078750 704 bp mRNA EST 18-OCT-1999
 DEFINITION AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone E0723_42, mRNA sequence.

ACCESSION AU078750
 VERSION AU078750.1 GI:6062509
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1 (bases 1 to 704)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from panicle at flowering stage
 Unpublished (1996)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = "RGP".

FEATURES
 source

1..704
 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"

BASE COUNT 203 a 167 c 142 g 191 t 1 others
 ORIGIN
 /clone="E0723_42"
 /clone_lib="Rice panicle at flowering stage"
 /dev_stage="flowering stage"
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

Query Match 64.0%; Score 16; DB 107; Length 704;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttcg 17
 |||||

Db 346 CATGCCAGCATTTCG 331

RESULT 6
 AZ578088/c 258 bp DNA GSS 08-DEC-2000
 LOCUS
 DEFINITION 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
 sp. NGR234 genomic clone 18f02, DNA sequence.

ACCESSION AZ578088
 VERSION AZ578088.1 GI:11605015
 KEYWORDS GSS.

SOURCE Rhizobium sp. NGR234.
 ORGANISM Rhizobium sp. NGR234
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 258)
 AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
 TITLE Genetic snapshots of the Rhizobium species NGR234 genome
 JOURNAL Genome Biology, vol 1 (6), 0014.1-0014.7 (2000)
 COMMENT Contact: Virginie Viprey
 Laboratoire de Biologie Moleculaire des Plantes Superieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.viprey@bsrc.ac.uk
 Class: shotgun.

FEATURES
 source
 1..258
 Location/Qualifiers
 /organism="Rhizobium sp. NGR234"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="18f02"
 /clone_lib="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="Vector: M13; derivative strain of NGR234 cured of
 pNGR234a"

BASE COUNT 51 a 82 c 79 g 46 t
 ORIGIN

Query Match 60.0%; Score 15; DB 246; Length 258;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgccagcatttcga 18
 |||||

Db 175 TCGCCAGCATTTCGA 161

RESULT 7
 BE065518/c 338 bp mRNA EST 09-JUN-2000
 LOCUS
 DEFINITION RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE065518
 VERSION BE065518.1 GI:8410168
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 338)

REFERENCE
AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-BT0316-170
 200-014-b06&t3=2000-02-17&t4=1)

Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 338.

FEATURES
 Location/Qualifiers
 1..338
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0316"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the puc 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 76 a 100 c 66 g 96 t

ORIGIN

Query Match 60.0%; Score 15; DB 162; Length 338;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 |||||

Db 97 CATGCCAGCATTTC 83

RESULT 8

AV551712 341 bp mRNA EST 06-SEP-2000

LOCUS AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana

DEFINITION cDNA clone RZ13e05R 5', mRNA sequence.

ACCESSION AV551712

VERSION AV551712.1 GI:8723125

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 341)

Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries

```

JOURNAL      DNA Res. 7, 175-180 (2000)
MEDLINE      20363093
COMMENT      Contact: Erika Asamizu
              The First Laboratory for Plant Gene Research
              Kazusa DNA Research Institute
              Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
              Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ13e05R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT   73 a  85 c  53 g  130 t
ORIGIN
      1  .341
      2  tcgccagcatttcca 18
      3  |||||||||||||
      4  322 TCGCCAGCATTTCGA 336

Query Match      60.0%; Score 15; DB 31; Length 341;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4  tcgccagcatttcca 18
Db  322 TCGCCAGCATTTCGA 336

RESULT 9
LOCUS      H60965      393 bp      mRNA      EST      06-OCT-1995
DEFINITION yrr2c07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
H60965
ACCESSION  H60965.1 GI:1013797
VERSION    H60965
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 393)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1078
High quality sequence stops: 299
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1078 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 299.
Location/Qualifiers
1. .393
/organism="Homo sapiens"
/db_xref="GDB:3775159"
/db_xref="taxon:9606"
/clone="IMAGE:206028"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

FEATURES
source
/organism="Homo sapiens"
/db_xref="GDB:3775159"
/db_xref="taxon:9606"
/clone="IMAGE:206028"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACGGAAGAAATTAATAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT   88 a  112 c  75 g  114 t
ORIGIN
      1  .394
      2  catcgccagcatttc 16
      3  |||||||||||||
      4  118 CATCGCCAGCATTTC 104

Query Match      60.0%; Score 15; DB 158; Length 393;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2  catcgccagcatttc 16
Db  118 CATCGCCAGCATTTC 104

RESULT 10
LOCUS      T86053/c      394 bp      mRNA      EST      17-MAR-1995
DEFINITION yd62q05.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
T86053
ACCESSION  T86053
VERSION    T86053
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1146
High quality sequence stops: 180
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1146 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 180.
Location/Qualifiers
1. .394
/organism="Homo sapiens"
/db_xref="GDB:468473"
/db_xref="taxon:9606"
/clone="IMAGE:112856"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACGGAAGAAATTAATAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

```

BASE COUNT      86 a    98 c    82 g    122 t    6 others
ORIGIN

Query Match      60.0%; Score 15; DB 189; Length 394;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  2 catgccagcatttc 16
    |||||
Db  183 CATGCCAGCATTTTC 169

RESULT 11
W23358
LOCUS      394 bp      mRNA      12-DEC-1996
DEFINITION SWAMCA507SK Brugia malayi adult male cDNA (SAW94NL-BmAM) Brugia
            malayi cDNA clone SWAMCA507 5', mRNA sequence.
ACCESSION  W23358
VERSION    W23358.1 GI:1300212
KEYWORDS   EST.
SOURCE     Brugia malayi.
ORGANISM   Brugia malayi.
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia.
REFERENCE  1 (bases 1 to 394)
AUTHORS   Williams, S.A.
TITLE     Genes expressed in adult males of Brugia malayi
JOURNAL   Unpublished (1995)
COMMENT   Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
Seq primer: pBluescript SK.
            Location/Qualifiers
                1..394
                /organism="Brugia malayi"
                /strain="RGS Labs"
                /db_xref="taxon:6279"
                /clone="SWAMCA507"
                /lab_host="XLI-Blue MRF"
                /note="vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
                Xho I; Lymphatic filarial nematode parasite of humans.
                mRNA was prepared from adult males of Brugia malayi
                isolated from birds and converted to double stranded cDNA
                using reverse transcriptase and oligo(dT) followed by
                RNase H and DNAPol I. The library had 4.6 x 10E6
                independent recombinants and average insert size was 800
                base pairs. The library was constructed by Noelle Ling.
                The library is available from Dr. S.A. Williams, email
                genome@smith.edu."
BASE COUNT      151 a    60 c    78 g    103 t    2 others
ORIGIN

Query Match      60.0%; Score 15; DB 190; Length 394;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  6 gccagcatttcgac 20
    |||||
Db  32 GCCAGCATTTTCGATC 46

RESULT 12
BE044431
LOCUS      399 bp      mRNA      08-JUN-2000
DEFINITION ho45d04.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
            IMAGE:1910913 3', mRNA sequence.
ACCESSION  BE044431
VERSION    BE044431.1 GI:8361484
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 414)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.

```

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 398.

FEATURES

source
 1. .414
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1910913"
 /clone_lib="NCI_CGAP_Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 122 a 82 c 109 g 101 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 18; Length 414;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 |||||
 Db 353 CATGCCAGCATTTC 367

RESULT 14
 AQ027122 435 bp DNA GSS 30-JUN-1998
 LOCUS CIT-HSP-2323H13.TF CIT-HSP Homo sapiens genomic clone 2323H13, DNA
 DEFINITION
 ACCESSION AQ027122
 VERSION AQ027122.1 GI:3267344
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 435)
 Adams.M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)

TITLE Unpublished (1998)
 JOURNAL Contact: Mark Adams
 COMMENT Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadamsetigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source
 1. .435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2323H13"
 /clone_lib="CIT-HSP"
 /sex="Male"

/cell_type="Sperm"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII"

BASE COUNT 140 a 97 c 77 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 223; Length 435;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 |||||
 Db 178 CATGCCAGCATTTC 192

RESULT 15
 H69200/c 437 bp mRNA EST 19-OCT-1995
 LOCUS Yr95f10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 DEFINITION IMAGE:213067 5', mRNA sequence.

ACCESSION H69200
 VERSION H69200.1 GI:1030526
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 437)

REFERENCE Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
 Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
 ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
 ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
 Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
 Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 1098
 High quality sequence stops: 300
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1098 Std Error: 0.00

Seq primer: M13RP1
 High quality sequence stop: 300.
 Location/Qualifiers
 1. .437

FEATURES

source
 1. .437
 /organism="Homo sapiens"
 /db_xref="GDB:3778371"
 /db_xref="taxon:9606"
 /clone="IMAGE:213067"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AACTGGAAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 97 a 113 c 89 g 134 t 4 others

BASE COUNT
 ORIGIN

AV722315/C

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AV722315 445 bp mRNA EST 16-OCT-2000
 AV722315 HTB Homo sapiens cDNA clone HTBAG01 5', mRNA sequence.
 AV722315
 AV722315.1 GI:10824678
 EST.
 human.

REFERENCE
AUTHORS

Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 445)

TITLE
JOURNAL
COMMENT

Li, N., Qian, B., and Zhang, J., 1997. *Cell*, 88, 103-110.
 Lu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
 Chen, J., Chen, Z., and Han, Z., 2000. Homo sapiens cDNA HTB clones
 Unpublished (2000)
 Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1129 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 364.

FEATURES

Location/Qualifiers
 1. 466
 /organism="Homo sapiens"
 /db_xref="GDB:3755192"
 /db_xref="taxon:9606"
 /clone="IMAGE:469190"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(gt) primer [5',
 AACTGGAGAAATTCGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."
 114 a 124 c 100 g 128 t

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 466;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16

Db 263 CATCGCCAGCATTC 249

RESULT 19

H60979/c
 LOCUS H60979 475 bp mRNA EST 06-OCT-1995
 DEFINITION YR22f01.rl Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:206041 5', mRNA sequence.

ACCESSION H60979

VERSION H60979.1 GI:1013811

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 475)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,

Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston

,R., Williamson,A., Wohlmann,P. and Wilson,R.

THE WashU-Merck EST Project

Unpublished (1995)

Contact: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1101

High quality sequence stops: 339

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1101 Std Error: 0.00

Seq primer: M13RP1

FEATURES

source
 Location/Qualifiers
 1. 475
 /organism="Homo sapiens"
 /db_xref="GDB:3775172"
 /db_xref="taxon:9606"
 /clone="IMAGE:206041"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(gt) primer
 [5' AACTGGAGAAATTAAGATCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 109 a 123 c 95 g 143 t 5 others

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 475;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16

Db 119 CATCGCCAGCATTC 105

RESULT 20

AI692369
 LOCUS AI692369 491 bp mRNA EST 17-DEC-1999
 DEFINITION wd63f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3',
 mRNA sequence.

ACCESSION AI692369

VERSION AI692369.1 GI:4969709

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 491)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 589 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 471.

Location/Qualifiers

1. 491

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2336283"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

151 a 94 c 129 g 116 t 1 others

BASE COUNT
ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 491;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16

|||||

Db 344 CATGCCAGCATTTC 358

RESULT 21

BF211097/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..491

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4047120"

/clone_lib="NIH_MGC 54"

/tissue_type="from chronic myelogenous leukemia"

/note="Organ: bone marrow; Vector: pONR-LIB (Clontech); Site_1: SfII (ggcgctgccc); Site_2: SfII (ggcgatattggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCGACATG-dh(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

123 a 128 c 119 g 121 t

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 146; Length 491;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgattcc 21

|||||

Db 475 CCAGCATTTCGATCC 461

RESULT 23

T84279/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1995)

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16

|||||

Db 305 CATGCCAGCATTTC 291

RESULT 22

AZ924681/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..492

/organism="Saccharomyces mikatae"

/strain="IFO 1815"

/db_xref="taxon:114525"

/clone="4906.id55a24.s1"

/clone_lib="Saccharomyces mikatae IFO 1815"

/note="Random genomic sequence"

149 a 83 c 91 g 169 t

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 251; Length 492;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgattcc 21

|||||

Db 475 CCAGCATTTCGATCC 461

RESULT 23

T84279/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1995)

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16

|||||

Db 305 CATGCCAGCATTTC 291

RESULT 22

AZ924681/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..492

/organism="Saccharomyces mikatae"

/strain="IFO 1815"

/db_xref="taxon:114525"

/clone="4906.id55a24.s1"

/clone_lib="Saccharomyces mikatae IFO 1815"

/note="Random genomic sequence"

149 a 83 c 91 g 169 t

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 251; Length 492;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgattcc 21

|||||

Db 475 CCAGCATTTCGATCC 461

RESULT 23

T84279/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1995)

Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 537.
 Location/Qualifiers
 1..539
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="H0353"
 /dev_stage="Adult"
 /note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 124 a 143 c 128 g 143 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 170; Length 539;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 catcgccagcatttc 16
 ||||||||||||
 Db 400 CATGCCAGCATTC 386

RESULT 26
 AW588900/c
 LOCUS AW588900 547 bp mRNA EST 05-APR-2001
 DEFINITION ra06g05.y2 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 cDNA 5' similar to TR:055013 O55013 BET3. ;contains Alu repetitive
 element; , mRNA sequence.
 ACCESSION AW588900
 VERSION AW588900.1 GI:7275932
 KEYWORDS EST.
 SOURCE southern root-knot nematode.
 ORGANISM Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
 REFERENCE 1 (bases 1 to 547)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
 Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisshvili,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
 ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
 Wilson,R.

TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Uma Rao and David Bird
 (david.bird@ucsu.edu) at North Carolina State University. DNA
 Sequencing by: Washington University Genome Sequencing Center St.
 Louis.
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 519.
 Location/Qualifiers
 1..547
 /organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="XLORL"
 /note="Vector: ZAP express - pBKCW (Stratagene); Site_1:

FEATURES
 Source
 1..547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="H0353"
 /dev_stage="Adult"
 /note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 124 a 143 c 128 g 143 t 1 others
 ORIGIN

Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 537.
 Location/Qualifiers
 1..185
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="H0353"
 /dev_stage="Adult"
 /note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 179 a 76 c 106 g 185 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 118; Length 547;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 catcgccagcatttc 16
 ||||||||||||
 Db 296 CATGCCAGCATTC 282

RESULT 27
 AA405993/c
 LOCUS AA405993 581 bp mRNA EST 09-NOV-1997
 DEFINITION zu66f09.r1 Soares_testis_NHR Homo sapiens cDNA clone IMAGE:742985
 5' similar to contains MER29.b3 MER29 repetitive element ; , mRNA
 sequence.
 ACCESSION AA405993
 VERSION AA405993.1 GI:2063976
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 581)
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 997 Std Error: 0.00
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 498.

FEATURES
 Source
 1..581
 /organism="Homo sapiens"
 /db_xref="GDB:5930030"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:742985"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pf73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pf7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 153 a 171 c 131 g 126 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 7; Length 581;

```

Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16
|||||
Db 81 CATCGCCAGCATTTC 67

RESULT 28
BE557935/c
LOCUS
DEFINITION
fl15d06.y1 zebrafish Research Genetics C32 fin Danio rerio cDNA 5'
similar to TR:O18973 O18973 RAB5 GDP/GTP EXCHANGE FACTOR, RABEX5.
; mRNA sequence.
BE557935
ACCESSION
VERSION BE557935.1 GI:9822425
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS Clark, M., Johnson, S. L., Lebrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.,
and Wilson, R.
WashU Zebrafish EST Project 1998
TITLE Unpublished (1998)
JOURNAL
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA library preparation: Ning Wu. cDNA Library Arrayed by:
Research Genetics. DNA Sequencing by: Washington University Genome
Sequencing Center. Clone distribution: Research Genetics web
address: http://www.researchgenetics.com/
Seq primer: T3 ET from Amersham
High quality sequence stop: 390.
FEATURES
source
Location/Qualifiers
1..622
/db_xref="taxon:7955"
/clone_lib="Zebrafish Research Genetics C32 fin"
/tissue_type="Fin"
/lab_host="Geneflags (HS996, a phage-resistant isolate of
DH10B)"
/note="Vector: pT73D-Pac with a modified polylinker;
Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared
from zebrafish(C32) fin, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is non-normalized. Library was
constructed by Ning Wu. NOTE: Clones from this library are
only available thru Research Genetics (www.resgen.com)."
```

BASE COUNT 164 a 160 c 164 g 134 t

ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 622;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcccagcatttga 18
|||||

Db 241 TCGCCAGCATTTCGA 227

RESULT 29
BE583417/c
LOCUS
DEFINITION
11-6H-MY Psojaemya Phytophthora sojae cDNA, mRNA sequence.
BE583417
ACCESSION
VERSION BE583417.1 GI:9834366
SOURCE EST.
ORGANISM Phytophthora sojae.
Phytophthora sojae.
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 630)
Qutob, D., Hraber, P. T., Sobral, B. W. S. and Gilzen, M.
Comparative analysis of expressed sequences in Phytophthora sojae
Plant Physiol. 123 (1), 243-254 (2000)
20267956
COMMENT Contact: Gilzen M
Agriculture and Agri-Food Canada
1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gilzenm@agr.ca.
Location/Qualifiers
1..630
/organism="Phytophthora sojae"
/strain="race 2, strain P6497"
/db_xref="taxon:67593"
/clone_lib="Psojaemya"
/dev_stage="mycelium"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
cDNA library was constructed from polyA+ enriched mRNA
from mycelium grown in liquid medium for 3 weeks at 25 C
in the dark in a liquid synthetic medium of 2.4 g sucrose
, 0.27 g asparagine, 10 mg cholesterol, 10 mg ascorbate,
and 2 mg thiamine, per liter, plus salts and minerals.
Complementary DNA was synthesized from mRNA using an
XhoI-poly(dT) linker-primer. EcoRI adapters were ligated
to the blunt-ended cDNA fragments and the products were
digested with XhoI for directional cloning into lambda ZAP
Express vector. This lambda library was amplified once
using E. coli host strain XL1 Blue MRF'. Inserts were then
subcloned by mass excision using ExAssist helper phage for
conversion into phagemid vector pBK-CMV in E. coli host
strain XL0LR. Sequenced using T3 primer: 5' ATT AAC CCT
CAC TAA AGG GA 3'."

BASE COUNT 135 a 184 c 169 g 140 t 2 others

ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 630;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16
|||||

Db 16 CATCGCCAGCATTTC 2

RESULT 30
BF156891/c
LOCUS
DEFINITION
f158b12.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
3818062.5, similar to TR:Q9UJ41 Q9UJ41 RAB5 GDP/GTP EXCHANGE FACTOR
HOMOLOGUE. ; mRNA sequence.
BF156891
ACCESSION
VERSION BF156891.1 GI:11052079
SOURCE EST.
ORGANISM zebrafish.
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.

```

REFERENCE
AUTHORS      1 (bases 1 to 684)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
WashU zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 Et from AmerSham
High quality sequence stop: 501.

FEATURES
source
1..684
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site:1: DraIII (CACTGCTGTG);
Site:2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCCTGGCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGCTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGCG and 3' end
primer CGACCTGCACCTCGACGACA."

BASE COUNT      177 a 177 c 193 g 137 t
ORIGIN

Query Match      60.0%; Score 15; DB 145; Length 684;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 tcgccagcatttcga 18
|||||
Db      237 TCGCCAGCATTCGA 223

RESULT 31
BF144032/c
LOCUS      BF144032      1096 bp      mRNA      EST      24-OCT-2000
DEFINITION      601791288F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4022252 5',
mRNA sequence.
ACCESSION      BF144032
VERSION      BF144032.1 GI:10983072
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1096)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9278 row: o column: 21
High quality sequence stop: 630.

FEATURES
Location/Qualifiers
1..1096
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4022252"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; transgenic model WNT-1, expression driven by
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT      336 a 271 c 271 g 218 t
ORIGIN

Query Match      60.0%; Score 15; DB 145; Length 1096;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 gcatttcgatctcgc 24
|||||
Db      1076 GCATTTCGATCCTCG 1062

RESULT 32
W06020/c
LOCUS      W06020      195 bp      mRNA      EST      22-MAY-2000
DEFINITION      TgESTz87f10.r1 TgrH Tachyzoite cDNA Toxoplasma gondii cDNA clone
tgz87f10.r1 5', mRNA sequence.
ACCESSION      W06020
VERSION      W06020.1 GI:1278733
KEYWORDS      EST.
SOURCE      Toxoplasma gondii.
ORGANISM      Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 195)
Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A., Aslett,M.A.,
Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
Waterston,R.H. and Boothroyd,J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
Contact: Marra,M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.edu
David Sibley at toxo@ebsci.wustl.edu for further information
relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 148.

FEATURES
source
1..195
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"

```

/clone="tgzy87f10.xl"
 /clone_lib="TgrH Tachyzoite cDNA"
 /lab_host="XLI-Blue MRP"
 /note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
 Toxoplasma RH strain tachyzoites were grown in human
 foreskin fibroblast cultures in vitro. The library was
 constructed by K.L. Wan, Cambridge University. cDNAs were
 synthesized from polyA RNAs by oligo d(T) priming and
 directionally cloned into the EcoRI to XhoI sites of the
 Lambda zapII vector using the ZAP-cDNA synthesis kit
 (Stratagene). WARNING: the library contains a small
 percentage of cDNAs derived from the human host cells."
 59 a 45 c 62 g 27 t 2 others

BASE COUNT
 ORIGIN

Query Match 56.0%; Score 14; DB 189; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcatttcgacccctc 23
 |||||

Db 181 GCATTTCGATCCTC 168

RESULT 33

AZ922752 200 bp DNA GSS 22-MAR-2001
 LOCUS SLCot4A05 Sorghum bicolor SLCot Sorghum bicolor genomic, DNA

DEFINITION sequence.

ACCESSION AZ922752

VERSION AZ922752.1 GI:13432973

KEYWORDS GSS.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 200)

Peterson,D.G., Schulze,S.R., Lee,S.A., Sciarra,E.B., Nagel,A.,
 Tibbitts,D.C., Wessler,S.R. and Paterson,A.H.

Characterization of the Sorghum bicolor genome using DNA
 renaturation kinetics (cot analysis) and repetition-based cloning

Unpublished (2001)

Contact: Peterson DG

Plant Genome Mapping Laboratory

University of Georgia

Room 162, Riverbend Research Bldg., 110 Riverbend Rd., Athens, GA

30602 USA

Tel: 706-583-0167

Fax: 706-583-0160

Email: dgp@arches.uga.edu

Class: Hydroxyapatite-fractionated DNA.

Location/Qualifiers

1..200

/organism="Sorghum bicolor"

/cultivar="Brx623"

/db_xref="taxon:4558"

/clone_lib="Sorghum bicolor SLCot"

/tissue_type="leaves"

/dev_stage="seedling"

/note="Vector: pGEM-TA-Easy; A Cot analysis was performed
 for the sorghum genome. Based on the resulting Cot curve,
 hydroxyapatite chromatography was used to isolate
 'highly-repetitive' (HR), 'moderately-repetitive' (MR),
 and 'single/low-copy' (SL) sequence components from
 sheared genomic DNA. The three repetition-based DNA
 components were cloned into E. coli to produce HRCot,
 MRCot, and SLCot genomic libraries. Blotting and
 sequencing data indicates that each library is
 representative of the component from which it was derived.
 Putative ID listings given for sequences are based on
 comparison (blastn) with sequences in the NCBI Nr

Database. Only the primary match is given (all primary E
 values are < or =3d 1.00E-5). In no instance does a 'Cot
 clone' contain the complete sequence of its putative Nr
 match."

BASE COUNT 48 a 49 c 57 g 42 t 4 others
 ORIGIN

Query Match 56.0%; Score 14; DB 251; Length 200;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattt 15
 |||||

Db 91 CATGCCAGCATT 104

RESULT 34

AW065325/c

LOCUS AW065325

DEFINITION 614045H11.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,

ACCESSION mRNA sequence.

VERSION AW065325

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 203)

Walbot,V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 614045 row: H column: 11.

Location/Qualifiers

1..203

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="614 - root cDNA library from Walbot Lab"

/tissue_type="root"

/dev_stage="3-4 days old"

/lab_host="XLOLR"

/note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
 EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"

BASE COUNT 46 a 52 c 51 g 54 t
 ORIGIN

Query Match 56.0%; Score 14; DB 111; Length 203;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atgccagcatttc 16
 |||||

Db 63 ATGCCAGCATTTC 50

RESULT 35

BG589383/c

LOCUS BG589383

DEFINITION EST497225 P. infestans-challenged leaf Solanum tuberosum cDNA clone

RPL11H7 5' sequence, mRNA sequence.

BG589383 203 bp mRNA EST 12-APR-2001

EST497225 P. infestans-challenged leaf Solanum tuberosum cDNA clone

RPL11H7 5' sequence, mRNA sequence.

```

ACCESSION      BG589383
VERSION        BG589383.1  GI:13607523
KEYWORDS       EST.
SOURCE         SOLANUM
ORGANISM       Solanum tuberosum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE      1 (bases 1 to 203)
AUTHORS        Zhang,P., Hernandez,M., Torndvist,C.-E, Wirtz,U., Loukolanov,A.,
               Rangell,P., Haberlandt,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell
               ,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
               Generation of ESTs from Potato Leaves Challenged with Phytophthora
               infestans, Incompatible Reaction
               Unpublished (2000)
JOURNAL        The Institute for Genomic Research
COMMENT        For clone info: please contact Research Genetics, Libraries
               Division tel 1-800-711-6195, email cdna@resgen.com
               Seq primer: M3F-R.
FEATURES       Location/Qualifiers
               1..203
               /organism="Solanum tuberosum"
               /cultivar="Kennebec"
               /db_xref="taxon:4113"
               /clone="BP11H7"
               /clone_lib="P. infestans-challenged leaf"
               /tissue_type="leaf"
               /dev_stage="6 week old"
               /lab_host="SOLR"
               /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
               XhoI; Whole plants were challenged with 450,000
               sporangia/ml P. infestans US-1(US 940501) in Biotron
               (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
               5, 12, and 24 hours post-challenge and frozen in liquid
               nitrogen immediately upon removal. Kennebec plants showed
               no signs of HR. Katahdin plants (susceptible to P.
               infestans US-1) were used as controls and showed
               infection. NOTE: We cannot exclude the possibility that
               this sequence is actually derived from Phytophthora rather
               than potato."
BASE COUNT    67 a 68 g 30 t
ORIGIN
Query Match   56.0%; Score 14; DB 155; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcacgacatttcg 17
Db 93 TCGCCAGCATTCG 80

RESULT 36
BG558891
LOCUS         RHI22_57_H09.bl_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
DEFINITION   sequence.
ACCESSION    BG558891
VERSION      BG558891.1  GI:13587889
KEYWORDS     EST.
SOURCE       Sorghum propinquum.
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
               clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE    1 (bases 1 to 205)
AUTHORS      Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
               ,L.H.
               An EST database from Sorghum: Sorghum propinquum rhizomes
               Unpublished (2000)
               Contact: Cordonnier-Pratt MM

BASE COUNT    205 bp mRNA EST 10-APR-2001
ORIGIN
Query Match   56.0%; Score 14; DB 155; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcacgacatttcg 17
Db 93 TCGCCAGCATTCG 80

RESULT 36
BG558891
LOCUS         RHI22_57_H09.bl_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
DEFINITION   sequence.
ACCESSION    BG558891
VERSION      BG558891.1  GI:13587889
KEYWORDS     EST.
SOURCE       Sorghum propinquum.
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
               clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE    1 (bases 1 to 205)
AUTHORS      Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
               ,L.H.
               An EST database from Sorghum: Sorghum propinquum rhizomes
               Unpublished (2000)
               Contact: Cordonnier-Pratt MM

BASE COUNT    205 bp mRNA EST 10-APR-2001
ORIGIN
Query Match   56.0%; Score 14; DB 155; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atgcacgacatttc 16
Db 141 ATGCCAGCATTC 154

RESULT 37
A0931496
LOCUS         A0931496.1  GI:6620510
DEFINITION   Mus musculus genomic clone RPCI-23-266P9,
               DNA sequence.
ACCESSION    A0931496
VERSION      A0931496.1  GI:6620510
KEYWORDS     GSS.
SOURCE       house mouse.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 211)
AUTHORS      Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
               ,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
               and Fraser,C.M.
               Mouse BAC End Sequences from Library RPCI-23
               Unpublished (1999)
               Contact: Shaying Zhao
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: szhao@tigr.org
               Clones are derived from the mouse BAC library RPCI-23. For BAC
               library availability, please contact Pieter de Jong
               (pieter@dejong.med.buffalo.edu). Clones may be purchased from
               BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
               or from Resea ch Genetics (info@resgen.com). BAC end page:
               http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
               Plate: 266 row: F column: 9
               Seq primer: SP6
               Class: BAC ends.
FEATURES       Location/Qualifiers
               1..211
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="RPCI-23-266P9"
               /clone_lib="RPCI-23"

```

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence start: 4
High quality sequence stop: 167
POLYA=No.

FEATURES Location/Qualifiers
 1..205
 /organism="Sorghum propinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RHIZ2)"
 /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
 Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda ZAP II.
 Clones to be sequenced were prepared by mass excision."
BASE COUNT 60 a 45 c 52 g 48 t
ORIGIN

Query Match 56.0%; Score 14; DB 155; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 atgcacgacatttc 16
Db 141 ATGCCAGCATTC 154

RESULT 37
A0931496
LOCUS A0931496.1 GI:6620510
DEFINITION Mus musculus genomic clone RPCI-23-266P9,
 DNA sequence.
ACCESSION A0931496
VERSION A0931496.1 GI:6620510
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 211)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
 ,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 266 row: F column: 9
 Seq primer: SP6
 Class: BAC ends.

FEATURES Location/Qualifiers
 1..211
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-266P9"
 /clone_lib="RPCI-23"

```

/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
```

BASE COUNT 82 a 38 c 28 g 62 t 1 others

ORIGIN

Query Match 56.0%; Score 14; DB 235; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gccagcattcgat 19
|||||

Db 159 GCCAGCATTCGAT 172

RESULT 38
A1329089/c
LOCUS 231 bp mRNA EST 28-DEC-1998
DEFINITION a9l2ne.r1 Neurospora crassa evening cDNA library Neurospora crassa
ACCESSION CDNA clone a9l2ne 3', mRNA sequence.
VERSION A1329089
KEYWORDS EST.
SOURCE A1329089.1 GI:4065648
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 231)
Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
Two Neurospora crassa EST Databases
Unpublished (1998)
JOURNAL
COMMENT Other ESTs: a9l2ne.fl
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Reverse Primer
High quality sequence stop: 179.

FEATURES
source
1..231
/organism="Neurospora crassa"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="a9l2ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/note="Vector: pBlueScript SK-; Site_1: XbaI; Site_2:
ECORI; See: Bell-Perdersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT 50 a 71 c 39 g 71 t

ORIGIN

Query Match 56.0%; Score 14; DB 18; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccatgcagcatt 14
|||||

Db 57 GCATCGCCAGCATT 44

RESULT 39
A1902416/c
LOCUS 242 bp mRNA EST 30-MAR-2000
DEFINITION CM-BT006-050299-125 BT006 Homo sapiens cDNA, mRNA sequence.
ACCESSION A1902416
VERSION A1902416.1 GI:6492803
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 242)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?tl=CM&t2=CM-BT006-125.html
&t3=050299&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers
1..242
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT006"
/sex="female"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
7716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 61 a 59 c 71 g 35 t 16 others

ORIGIN

Query Match 56.0%; Score 14; DB 103; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gccacatttcgat 19
|||||

Db 219 GCCAGCATTCGAT 206

RESULT 40
BB090429
LOCUS 251 bp mRNA EST 23-JUN-2000
DEFINITION BB090429 RIKEN full-length enriched, 12 days embryo, embryonic body
between diaphragm region and neck Mus musculus cDNA clone
9430026F22 3', similar to AL080159 Homo sapiens mRNA; cDNA
DKFZp434M154 (from clone DKFZp434M154);, mRNA sequence.
ACCESSION BB090429

ORIGIN

Query Match 56.0%; Score 14; DB 115; Length 271;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgatc 20
 |||||

Db 264 CCAGCATTTCGATC 251

RESULT 42

AA165335
 LOCUS
 DEFINITION
 IMAGE:633034 5', mRNA sequence. 18-DEC-1996
 ACCESSION
 AA165335
 VERSION
 AA165335.1 GI:1741484
 KEYWORDS
 EST.
 SOURCE
 human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 286)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chisoe, S., Dietrich, N., DuBuque, T., Faville, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 B., Mullis, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

TITLE

Generation and analysis of 280,000 human expressed sequence tags

JOURNAL

Washington University School of Medicine

MEDLINE

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

COMMENT

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 178.

FEATURES

source

Location/Qualifiers
 1..286
 /organism="Homo sapiens"
 /db_xref="GDB:5183118"
 /db_xref="taxon:9606"
 /clone="IMAGE:633034"
 /clone_lib="Stratagene hNT neuron (#937233)"
 /dev_stage="hNT neurons"
 /lab_host="SOLR (kanamycin resistant)"
 /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt.
 Differentiated, post mitotic hNT neurons. Average insert
 size: 1.5 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'
 GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTTTTTT 3'
 55 a 120 c 66 g 38 t 7 others

BASE COUNT

ORIGIN

Query Match 56.0%; Score 14; DB 3; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcacgcgcgcatt 14

Db 241 GCATCGCGACGATT 254

RESULT 43

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989 311 bp mRNA EST 30-JUN-2000
 BB216989 RIKEN full-length enriched, adult male aorta and vein Mus
 musculus cDNA clone A530040F06 3', mRNA sequence.
 BB216989
 BB216989.1 GI:8881942
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 311)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
 T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp/
 URL: http://genome.rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Akiyama, J., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
 1..311
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A530040F06"
 /clone_lib="RIKEN full-length enriched, adult male aorta
 and vein"
 /sex="male"
 /tissue_type="aorta and vein"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization

